Understanding heterogeneity of gene expression in the innate immune system through mathematical modelling

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2023

Nissrin Alachkar

School of Biological Sciences
Division of Infection, Immunity and Respiratory Medicine

Table of Contents

Word count: 56,789

List of Figures

List of Abbreviations

smFISH: Single molecule fluorescence *in situ* hybridisation

scRNA-seq: Single-cell RNA sequencing

cfp: Cyan fluorescent protein

yfp: Yellow fluorescent protein

gfp: Green fluorescent protein

RNAP: RNA polymerase

TF: Transcription factor

TBP: TATA box-binding protein

OPN: occupied proximal nucleosome

$\boldsymbol{b}$**:** Burst size

$\boldsymbol{f}$**:** Frequency

MS2: Mass spectrometry

TNF-α: Tumour Necrosis Factor α

TSA: Trichostatin A

NF-Y: Nuclear transcription factor Y

TLR: Toll-like receptor

PRRs: Pattern recognition receptors

NK: Natural killer cells

PAMPs: Pathogen-associated molecular patterns

dsRNA: Double-stranded RNA

PIC: Polyinosinic-polycytidylic acid

LPS: Lipopolysaccharides

TIR: Toll/interleukin-1 receptor

TRIF: TIR-domain-containing adaptor protein inducing interferon-β

MyD88: Myeloid differentiation primary response gene 88

NF-κB: Nuclear factor kappa-light-chain-enhancer of activated B cells

IRF: Interferon regulatory factor

IL: Interleukins

IFN-β: Type I interferon-beta

TS: Transcription sites

FACS: Fluorescent activated cell sorting

ODE: Ordinary differential equation

CME: Chemical master equation

FSP: Finite state projection

MCMC: Markov chain Monte Carlo

MH: Metropolis-Hastings

MLE: Maximum likelihood estimation

AIC: Akaike information Criterion

BIC: Bayesian Information Criterion

DIC: Deviance information criterion

IS-HME: Importance Sampler of the Harmonic Mean Estimator

Abstract

Mammalian gene transcription is rigorously regulated through various complex mechanisms to ensure accuracy and prevent errors in the synthesis of RNA molecules. Paradoxically, transcriptional cell-to-cell variability, despite genetic homogeneity, has been broadly documented, including in the innate immune system where precise responses are crucial upon encountering pathogens. The stochastic nature of transcription is believed to be a driver of this variability; the transcriptional process of most genes involves random transitions between inactive and active gene states, leading to the production of messenger RNA (mRNA) in a burst-like manner, giving rise to inherent heterogeneity in gene expression at the single-cell level. While this phenomenon has been studied for decades, it remains unclear whether, and how, single-cell variability in the innate immune system is controlled in response to different environmental conditions.

Combining data analysis, statistical inference, and state-of-the art mathematical modelling with data from various wet lab techniques, this thesis presents interdisciplinary research on characterising cellular variability in the innate immune Toll-like receptor system and provides new understanding of the underlying control mechanisms. The first chapter introduces the notion of gene expression heterogeneity with an overview of the existing relevant literature and discusses various approaches from both the biological and mathematical fields that have been or can be potentially employed to study this phenomenon. Chapter two focuses on analysis and mathematical modelling of single molecule fluorescence in situ hybridisation count data of inducible immune genes. Gene-specific linear mean-variance relationships of mRNA transcript counts across a range of immune conditions, and their corresponding bursting characteristics, are established. Chapter three validates the linear constraints, and their underlying transcriptional bursting modulation, globally and demonstrates, through stochastic modelling of single-cell RNA-seq counts of 96 immune genes, an association between high variability levels and increased complexity of transcriptional regulation. In addition, evolutionary differences in response variability across several species are characterised. Chapter four provides evidence that heterogenous single cell innate immune responses are in part imprinted over multiple cell divisions. Overall, this thesis offers novel tools and findings that take us a step forward in understanding cell-to-cell variability in the innate immune system with broader implications for other biological systems.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university of other institute of learning.

Copyright Statement

Dedication

This work would not have been accomplished without the immense support I had from my family and friends. My heart is filled with love, appreciation, and gratitude to each one of you.

To my dad, my inspiration, thank you for the countless hours you spent listening to my attempts at explaining the work, for always guiding me towards finding solutions for the many problems I faced, and for discussing and challenging me with intriguing scientific ideas that kept me stimulated and motivated to work harder. Thank you for being a mentor, teaching me how to think logically and for introducing me to the beauty of research. Most importantly, thank you for believing in me from day one and encouraging me to take this path.

To my mum, the most loving and caring, thank you for taking care of the smallest details concerning my well-being, for being understanding and patient through my most stressful times, and for never failing to lift me up when I needed it. You created an environment full of warmth and comfort for me while writing my thesis, and for that I am grateful. From you I learnt how to be persistent and patient, two of the most needed qualities to finish a PhD.

To my partner, Tom, you have been my peace and calmness throughout this journey. Thank you for standing by my side every step of the journey, for your patience, and for always believing in me and motivating me to keep going.

To my siblings, Sawsan and Ahmad, and my uncle Mustafa, thank you for always checking on me, for being involved in the little details, and for your constant words of encouragement.

To my friends, thank you for always being there for me, for the great times that helped my sanity, and for your support with the nicest gestures. I will forever cherish the memories made with you.

Finally, I dedicate the success of this work to a beautiful soul that left us during my PhD, to my dear aunt Malack. You always supported me and were the first to celebrate my successes. The memory of your radiant smile has been a source of strength during my toughest times of this journey. You continue to be an inspiration that guides me in everything I do.

Acknowledgements

I would like to express my sincere gratitude to my two supervisors Dr. Pawel Paszek and Dr. Mark Muldoon. I am deeply grateful for their continuous guidance, patience, and support throughout my doctoral journey. Their complementary expertise and knowledge together with their enthusiasm created the best environment for me to dive into this multidisciplinary research experience after finishing my master's in mathematics. Their constructive feedback has consistently elevated the quality of my work and pushed me to deliver my best. I learnt a lot of invaluable skills from each of them that nurtured my academic and personal growth and will no doubt help me in my next career step.

My appreciation is also to the QBB PhD programme, led by Prof. Nancy Papalopulu, for granting me this incredible opportunity. I consider myself lucky to be part of such an intellectual and scientific community. The regular meetings and intriguing scientific discussions with the QBB cohort have been stimulating and inspiring. I specially thank Alex Koch for his support and friendship. From assisting with experiments to helping me navigate last-minute technical issues, his genuine help has been invaluable.

I would also like to thank the Wellcome Trust for their generous financial support.

Many thanks go to Dr. Josephine Morgan and Dr. James Bagnall for providing the smFISH, and RNA-seq data presented in this thesis and for their expert advice and guidance. Thanks also go to Abhyduai Singh whose collaboration provided inspiration for the last chapter of this thesis. I would also like to thank David spiller for showing me how to use a microscope and helping me troubleshooting any imaging issues.

Additionally, I would like to extend my appreciation to my final year officemates in the Maths department; Jake Harris, Miko Widaswki, Tom Hitchen and Adrien Weihs for the many interesting scientific and non-scientific debates. You made my final year much more enjoyable.

Rationale

The thesis is presented in the alternative (journal) format consisting of two peer-reviewed published articles and one unpublished manuscript, and is organised as follows:

- Chapter two contains the article published in Cell Systems; *Gene-specific linear trends constrain transcriptional variability of the Toll-like receptor signaling* (Bagnall et al., 2020).
    - This article demonstrates that the variance of TLR-induced gene expression is linearly constrained by mean mRNA response across a range of stimulations. This phenomenon is underlined by reciprocal relationship between burst size and frequency. Variability of two cytokines *TNF*α and *IL*1β is analysed in details and stochastic models of gene expression are presented to capture their variability. Experimental approaches used include smFISH and scRNA-seq.
    - For this publication, I developed the mathematical methodology for analysis of transcriptional bursting and the mathematical models of *TNF*α and *IL*1β under the supervision of Dr Mark Muldoon and Dr Pawel Paszek. I performed the mathematical calculations, computational analysis and modelling presented in the paper. I also generated figures (Figure 4, S6, S14 and S15).
- Chapter three contains the article published in Frontiers in Molecular Biosciences; *Variability of the innate immune response is globally constrained by transcriptional bursting* (Alachkar et al., 2023).
    - This article presents a genome- and species-wide analysis of transcriptional variability in the innate immune system. It confirms the linear mean-variance constraints as a global control feature of cell-to-cell variability in TLR system across species and introduces analytically predicted burst size and frequency modulations driving the constraints. A new stochastic model of gene expression is introduced and relationship between model complexity and the level of transcriptional variability is established.
    - For this publication, I performed the mathematical calculations, computational and data analyses, generated the figures, and assisted with writing the manuscript alongside Dr Pawel Paszek and Dr Mark Muldoon.

- Chapter four contains the unpublished article; *Single-cell gene expression patterns of the Toll-like receptor signalling are heritable traits.*
    - This chapter demonstrates that the heterogenous single-cell TLR responses are heritable across multiple cell divisions. These findings challenge the conventional stochastic modelling of gene expression and call for a re-evaluation of the underlying assumptions.
    - For this publication, I carried out the mathematical calculations and data analyses, generated the figures, carried out the immunostaining experiment of Cd36 and assisted with writing the manuscript alongside Dr Pawel Paszek.

Chapter 1 Introduction

1.1 Cellular variability in gene expression

Heterogeneity of a single trait in a population of cells or individuals, also referred to as 'non-genetic individuality' believed to cause phenotype variations, is hardly a new concept in biology (Grote et al., 2015; Spudich & Koshland, 1976). Gene expression heterogeneity is a great evident example of this concept. Despite the tight regulation of the gene expression processes, cell-to-cell variability in the mRNA and protein levels across a genetically homogeneous population has been ubiquitously observed across different species, from viruses and bacteria to mammals (Balaban et al., 2004; Loewer & Lahav, 2011; Raj et al., 2006; Schulte & Andino, 2014; Taniguchi et al., 2010), and within different tissues (Sturm et al., 2021), raising many fundamental biological questions of how and why. The work of this thesis explores the phenomenon of single-cell transcriptional heterogeneity and aims to understand how transcriptional variability is controlled during innate immune responses.

In this chapter, I review the current understanding of gene expression heterogeneity on a biological level through discussing its origins and biological significance for cellular systems, in particular the immune system, the role transcriptional bursting plays in producing this heterogeneity, as well as studies and methodologies capturing the phenomenon.

1.1.1 Origins and consequences of cellular variability in gene expression

Cell-to-cell variability in the number of mRNA transcripts and protein molecules in genetically identical population has been well documented, with the revolution of single-cell experimental techniques, in both prokaryotes and eukaryotes (Becskei et al., 2005; Blake et al., 2003a; Elowitz et al., 2002; Golding et al., 2005; Ozbudak et al., 2002). These studies demonstrated that the behaviour of individual cells may differ from the population average, a phenomenon that has been hidden for many years in cell population studies. In one of the earliest experimental studies to observe heterogeneity in gene expression, Novick and Weiner showed that synthesis of the enzyme β-galactosidase in individual cells of the bacterium *Escherichia coli (E. coli)* was variable (Novick & Weiner, 1957). Upon induction with low inducer concentrations, individual cells did not synthesise a common level of the enzyme but rather, a subset of bacteria produced β-galactosidase at full rate while the rest did not

make any. Thirty-three years later, Ko *et al.* introduced expression reporters in single-cells to examine the expression of a dose-sensitive glucocorticoid-responsive transgene encoding β-galactosidase, upon induction of different doses of glucocorticoid (M. S. Ko et al., 1990). Through this novel technique, they found that increasing the dose level of glucocorticoid resulted in higher number of cells producing β-galactosidase, but without a uniform increase in every cell's expression. In a following study, Ko explained the experimental results by producing computational simulations from a stochastic model of gene induction, where stochasticity is imagined to arise from the random timing of molecular collisions and dissociations between transcription factors and a gene copy (M. S. H. Ko, 1991).

These studies evoked the fundamental question of why seemingly identical cells exhibit variability in gene expression, thus paving the way for a new field of gene expression regulation research focused on studying and unravelling the underlying mechanisms of cellular heterogeneity, while incorporating multidisciplinary approaches. Nowadays, this heterogeneity has been observed in various biological systems, indicating that it is not a result of mere chance but rather a fundamental and regulated aspect of cellular processes.

Extrinsic versus Intrinsic heterogeneity

Observed heterogeneity of gene expression among a clonal (isogenic) population can be divided into two components - extrinsic and intrinsic - depending on its origins (Elowitz et al., 2002; Horsthemke et al., 1992; Schuster & Érdi, 1989; Swain et al., 2002). Extrinsic heterogeneity accounts for cell-to-cell variability resulting from non-uniform extrinsic factors (changes in global and cellular environment of a gene) affecting individual cells at different levels. For example, cells of a clonal population can be in different cell cycle stages (Rosenfeld et al., 2005), and can have different cell size (Kempe et al., 2015; Padovan-Merhar et al., 2015). Even cell culture medium has been shown to influence gene expression with serum-based medium increasing the response variability (G. Guo et al., 2016). Cell-to-cell differences in the amount or activity of regulatory proteins and polymerase molecules (which in turn can be a result of neighbouring cells activities and signalling) can also add to extrinsic heterogeneity by changing the rates of fundamental reactions that impact gene expression (Raser & O'Shea, 2004, 2005) as gene expression is controlled by their concentrations, states, and locations. Such variables influencing extrinsic heterogeneity are not always possible to

identify and quantify, however it is important to take these hidden factors into account when studying gene expression variability (S. Huang, 2009; Swain et al., 2002). In comparison, intrinsic heterogeneity refers to the temporal fluctuations of mRNA expression in an individual cell, and since these fluctuations are not synchronised between cells of the same population, they result in overall variability within the population. These fluctuations are believed to arise from two major factors. First, the stochastic nature of gene expression which conceals the random microscopic events that dictate the occurrence and sequence of biochemical reactions within a cell. This is attributed to the fluctuations in the low number of molecules involved in these reactions, a phenomenon known as the 'finite number effect' (Elowitz et al., 2002). Second, epigenetic factors, like DNA methylation and histone modification, that cause conformational changes to the gene influencing its accessibility and thereby its activity (Dey et al., 2015; Handy et al., 2011; Suter et al., 2011; Viñuelas et al., 2013). Both factors will be discussed in more details in **section 1.1.3**, and while I discuss them separately, it is important to keep in mind their dynamic interplay and the mutual influence they exert on each other (Beckman et al., 2021; Capp, 2021).

Elowitz *et al.* distinguishably quantified gene expression extrinsic and intrinsic components of heterogeneity by incorporating two distinguishable reporter genes, cyan fluorescent protein (cfp) and yellow fluorescent protein (yfp), into the genome of *E. coli* such that both genes were controlled by identical promoters (Elowitz et al., 2002). In the absence of intrinsic noise, the extrinsic component was then reflected in the equal fluctuations in expression of the two genes in an individual cell, but differences in expression levels from cell to cell. The intrinsic component was measured by the uncorrelated variations in the amount of cfp and yfp within individual cells. They found that both sources of noise can be significant depending on the transcription rate, regulatory dynamics, and genetic factors. Ozbudak *et al.* also provided evidence of intrinsic noise in bacteria and demonstrated that the level of phenotypic variation in isogenic population can be regulated by epigenetic parameters (Ozbudak et al., 2002). By altering the level of induction and introducing mutations into the ribosomal binding site, they quantitatively measured, using flow cytometry, the changes in phenotypic noise characteristics as a result from varying the rates of transcription and translation of green fluorescent protein (gfp) in *Bacillus Subtilis* cells. Interestingly, they found that increased translational efficiency is the predominant source of increased protein noise which, in turn,

depends on the stochastic mechanisms of gene activation (Kierzek et al., 2001; McAdams & Arkin, 1997)

In mammalian cells, intrinsic noise, specifically the stochastic nature of gene expression, has been shown to be the major contributor to transcriptional heterogeneity (Raj et al., 2006), in addition to evidence supporting allelic imbalances (when two alleles of the same gene are expressed at different levels independently) due to possible epigenetic allelic differences (Pastinen et al., 2004; Wagner et al., 2010). Bahar Halpern *et al*., provided evidence of bursty expression in mammalian liver tissues contributing to noise, with genes that exhibit more burst-associated noise having longer mRNA lifetime to reduce overall temporal variability (Bahar Halpern et al., 2015). In contrary, a more recent study demonstrated that cell state differences are the responsible drivers for heterogeneity in gene expression (Foreman & Wollman, 2020).

Although it is important to understand the differences between the two components (extrinsic and intrinsic) that drive heterogeneity, it is, however, worth asking the question of whether it is reasonable to attempt to distinguish between them experimentally. Many studies attempt to "control" as many extrinsic factors as they can in order to decompose intrinsic heterogeneity. However, it could be argued that eliminating those hidden factors from the process would essentially remove the randomness factor of the process (Symmons & Raj, 2016). One way to look at it would be that fluctuations in extrinsic factors could also be stochastic but on a much slower timescale than that of intrinsic factors (Rosenfeld et al., 2005; Shahrezaei & Swain, 2008a). In addition, genes work as part of regulatory circuits, and intrinsic noise in one gene can be seen as an extrinsic noise factor in another downstream gene (Hooshangi et al., 2005). For that reason, some might argue that all intrinsic and extrinsic factors distinguished in the field emerge from intrinsic noise, with the difference that intrinsic noise is inherently initiated, while extrinsic noise is transmitted (Brown & Boeger, 2014).
The different nature but interplay of the extrinsic and intrinsic components also imposes the fundamental question of whether the heterogeneity trait is stochastic (non-deterministic) or pre-determined to some extent (Ansel et al., 2008). If the latter is the case, we would expect the trait to persist for at least several generations of clonal samples. In recent studies, it has been discovered that gene transcription in rare (few individuals within a population) cells can

last for longer timescale than multiple cell divisions, i.e., having cellular memory over a multi-generational timescale (Phillips et al., 2019; Shaffer et al., 2020). I explore this question further in chapter four.

1.1.2 Is transcriptional heterogeneity just unwanted noise?

Regardless of the origins of heterogeneity (intrinsic and extrinsic) described above, the measured level of variability in gene expression among genetically identical cells is also referred to as "noise" in gene expression. However, noise is often used to describe unwanted disturbance in a system, so is noise in gene expression a harmful error, that could be avoidable, or is it a beneficial evolutionary strategy adopted by living organisms? Variations in the expression level of a gene within genetically identical individual cells allow heterogeneous phenotypes in clonal populations. This can benefit a population when experiencing sudden fluctuations or switching between environments through containing subpopulations that are devoted to different behaviours (Kussell & Leibler, 2005; Lehner, 2008). The consequences of this bet hedging strategy on different biological processes have been observed including the development of bacterial antibiotic resistance (Maisonneuve & Gerdes, 2014; Verstraeten et al., 2015); *E. coli* continuously produces a small subpopulation that is antibiotic-resistant even in untreated growing cultures (Balaban et al., 2004). Induction of stress responses is another consequence of this strategy; studies showed that stress-related genes display higher levels of noise than those related to other biological functions (Newman et al., 2006; Silander et al., 2012), with increased noise only providing a significant selective advantage at high environmental stress levels (Blake et al., 2006). In higher eukaryotes, the functional importance of gene expression heterogeneity is still not obvious but some studies showed the effect of noise in triggering random cell-fate decisions during development (J.-Y. Chen et al., 2012; Magklara & Lomvardas, 2013; Spencer et al., 2009) leading to phenotypic diversity as a result, and allowing a wide range of possible cellular behaviours in homogeneous cell types; retinal mosaic (Wernet et al., 2006) and dynamic mosaic endothelial cell heterogeneity permitting adaptive homeostasis (Yuan et al., 2016) are examples. On the other hand, gene expression noise can be harmful in some cases, as it may limit the precision of cellular processes (Elowitz & Leibler, 2000).

With evidence supporting both sides of the argument, gene expression noise is believed to be not only a heritable genetic trait (Ansel et al., 2008), but also an evolvable one through natural selection, providing a higher chance of survival (Fraser & Kaern, 2009; Paszek et al., 2010; Stern et al., 2007). In fact, in a relatively recent study (Hagai et al., 2018), it has been shown that gene expression variability plays a role in shaping the innate immune system through evolution. They found that genes exhibiting high transcriptional evolutionary divergence in stimulus-response across species also displayed high cell-to-cell variability within a species. Those genes were classified to be cytokines and chemokines (involved in the activation of immune responses, discussed later). In comparison, the more conserved genes in their transcription across species had low variability among their individual cells and were related in functionality to immune response regulation. One explanation of this could be that cell-to-cell variability allows more freedom for immune genes with certain functionalities to undergo rapid evolution through evolutionarily mechanisms (e.g., to adapt to the evolution of pathogens) in comparison to genes that need more constrained evolution. Cytokines, for example, are among the fastest evolving genes (Antczak et al., 2022; Scapigliati et al., 2006).

Nevertheless, comparing the characteristics of variability across different species would be one way to provide a better understanding of the potential functionality and consequences of this trait.

1.1.3 Transcriptional bursting and epigenetic roles in variability of gene expression

Gene expression is one of the most fundamental cellular processes in all living organisms. Through gene expression genetic information encoded in DNA is transcribed into RNA and translated into proteins, determining in turn the functionality and fate of cells in an organism. With the advancements of single-cell experimental techniques and new technologies, it has been experimentally confirmed that transcription (DNA to RNA) occurs by random transitions between states of gene activity and inactivity (Bahar Halpern et al., 2015; Suter et al., 2011), resulting in periods of mRNA production with a timescale varying from few minutes up to multiple hours. The discontinuous nature of this phenomenon over time is referred to by the term “transcriptional bursting” (Tunnacliffe & Chubb, 2020). Hence, to avoid ambiguities, this term does not carry any implicit implications regarding a specific model or mechanism behind the process.

Integrating experimental methods with mathematical modelling showed that transcription in a bursty manner contributes to the expression noise (Blake et al., 2003a; Elowitz et al., 2002; Ozbudak et al., 2002; Raj et al., 2006; Raser & O'Shea, 2004). In mammalian cells, Raj *et al*. provided direct evidence of cell-to-cell variations in mRNA levels due to infrequent random gene transitions between active and inactive states (Raj et al., 2006).

The ordered assembly of RNA polymerases (RNAP), transcription factors (TF) together with their associated transcription mediator complexes binding to cis-regulatory elements (including promoters, enhancers, silencers, and insulators) of the DNA at a particular time following chromatin remodelling, which refers to the dynamic modification of chromatin structure (including histone modifications and nucleosome eviction), is the key for initiating and regulating transcription (Johnson et al., 2002) (**Figure 1.1**). Therefore, the interplay of these molecules constitutes key mechanisms underlying the burstiness nature of transcription. Studying how these molecular components of transcription contribute to transcriptional bursting and cell-to-cell variability is of a great interest in the field.

Promoter architecture (defined by the number, strength and position of transcription factor and RNAP binding sites) has been shown to affect cell-to-cell variability levels (Jones et al., 2014; Sanchez et al., 2011; Sharon et al., 2014). In *Saccharomyces cerevisiae,* it has been revealed that both the position and the number of repressor binding sites – namely operators – can influence cellular variability (Murphy et al., 2007). An increase in the number of the operators and a decrease in the distance of the operator site within a promoter to the TATA-box results in an increase in the noise level.

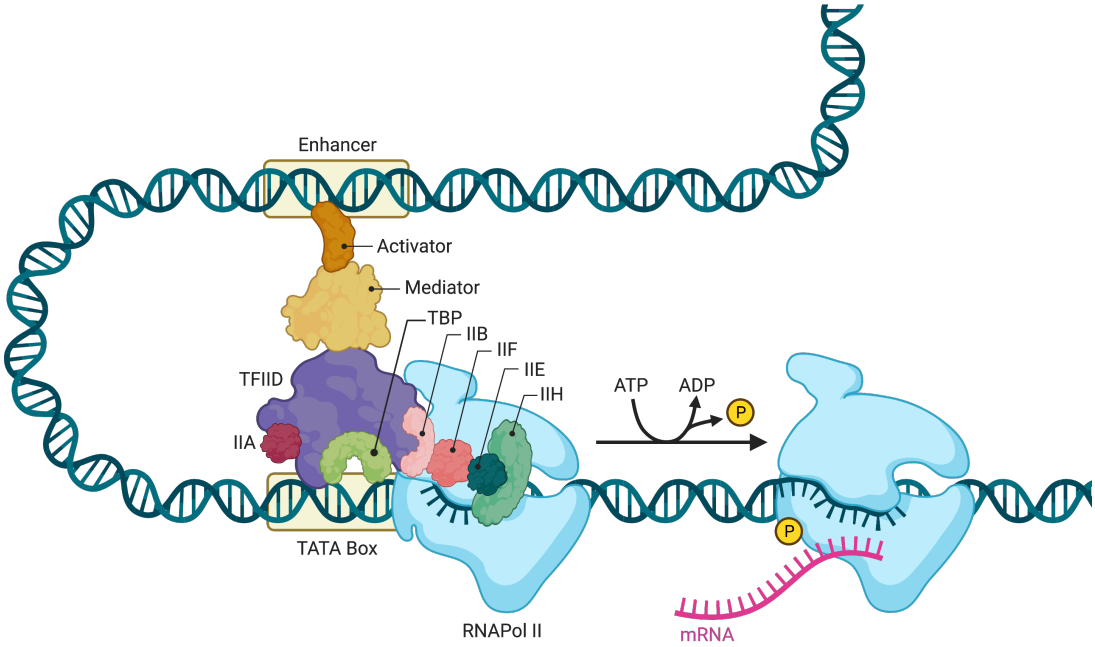


Figure 1.1: Schematic representation of transcriptional initiation. *Adapted from "Eukaryotic Gene Regulation - Transcriptional Initiation", by BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates. Shown is the transcription initiation complex: RNAPol II and general transcription factors (TBP, IIB, IIF, IIH, TFIID and IIA) attached to the TATA Box region of the promoter to activate transcription. Activator proteins bind to the enhancer part of the DNA to stabilise promoter binding through DNA looping and interactions with a mediator.*

Other established regulatory features of transcriptional variability associated with genes promoters' structure are CpG islands (DNA methylations regions) and TATA-boxes (transcription factors binding sites). Short CpG islands and the existence of a TATA-box motif (not all genes have them, some genes have TATA-like sequence (Rhee & Pugh, 2012)) in the core promoter are associated with increased transcriptional noise (Blake et al., 2003b; Morgan & Marioni, 2018; Ravarani et al., 2016). This effect of TATA-box presence can be indirectly related to the affinity and competition for TATA box-binding protein (TBP), which in turn can be affected by the specific co-activator complex (SAGA or TFIID) binding to it. Recruitment of TBP has a major role to play in regulating the output of gene expression in general (Kim & Iyer, 2004). Contrary to what the name suggests, TBP can bind to promoters with both TATA-box and TATA-like sequences (Rhee & Pugh, 2012). However, TATA-box promoters exhibit higher flexibility in recruiting TBP (Zhou et al., 2013). They show more competition with higher affinity for TBP which arises from their interplay with co-activators complex bound to the TATA-box (Cianfrocco et al., 2013; Ravarani et al., 2016). It has been shown that promoters containing a TATA-box tend to bind to SAGA, while promoters with a TATA-like sequence mostly recruit TFIID, with the former exhibiting higher noise levels

(Ravarani et al., 2016). The binding of TBP with TFIID is a more stable complex, leading to longer residence time of TBP at the TATA-like promoters, and hence by a consistent transcriptional output with low noise levels in comparisons to less stable TBP:SAGA complex (de Jonge et al., 2017; Ravarani et al., 2016).

Overall, the type of transcription binding sites, TBP, the corresponding co-activators complex assembled at the promoter and their complex interactions are all responsible, to some extent, for cell-to-cell variability.

Epigenetic processes (alterations of transcriptional potential of a cell without modifying the underlying DNA sequence (Waterland, 2006)), including DNA methylation and histone modification, play a crucial role in regulating gene expression and shaping mRNA output (Gibney & Nolan, 2010).

Differences in promoter nucleosome configuration in the same gene has been shown to contribute to the gene expression noise (Brown & Boeger, 2014). In yeast, promoter nucleosome occupancy was linked to transcriptional plasticity (flexibility of a gene to alter its expression in response to stimuli) (Tirosh & Barkai, 2008). Further, genes with occupied proximal-nucleosome (OPN), an established nucleosome occupancy promoter pattern, defined by the lack of a nucleosome-free region in which it prevents promoter accessibility to different proteins needed for transcription, exhibited high levels of expression noise (Field et al., 2008; Tirosh & Barkai, 2008).

Transcriptional bursting characteristics

The output of the resulting stochastic bursts during transcriptional activity is characterised by their size, defined as the number of mRNA produced per transcriptional burst event, and their frequency which is the number of bursts per unit time (Raj et al., 2006; Raj & van Oudenaarden, 2008; Suter et al., 2011), and these characteristics form a terminology framework to study gene expression variability as will be discussed.

Several studies presented work prompting the hypothesis that different genes display different bursting characteristics governed by gene-specific bursting mechanisms (Muramoto et al., 2012; Skinner et al., 2016; Suter et al., 2011). For instance, transcriptional burst

frequency in mammalian cells can vary from a burst every 30 minutes up to 10 hours, while burst size ranges from one to few hundreds mRNA molecules (Lionnet & Singer, 2012).
Mathematically, and under the condition that transcription is happening in short and infrequent bursts with respect to the timescale of mRNA (and sometimes protein) half-life, with large burst size, approximations for both burst size b and burst frequency f can be obtained in terms of the mean μ and variance σ^2 of the mRNA transcript numbers, with $\boldsymbol{b} = \frac{\boldsymbol{\sigma^2}}{\boldsymbol{\mu}}$ (i.e the Fano factor, which has been used to quantify cell-to-cell variability as being the noise strength), while $\boldsymbol{f} = \frac{\boldsymbol{\mu}}{\boldsymbol{b-1}}$ (Nicolas et al., 2017; So et al., 2011). These expressions capture departures from a non-bursty (Poissonian) regime where variance and mean are equal, for which burst size is equal to one and frequency is infinity. I refer to these throughout the thesis as moment estimators.

Studying how molecular mechanisms of transcription alter the bursting characteristics in the cell and whether these characteristics are global or gene-specific is one fundamental approach to refine our understanding of the regulation of transcriptional bursting and hence cell-to-cell variability in mRNA at single-cell level.
Over the years, studies using different experimental approaches, such as smFISH and MS2-GFP, integrated with mathematical approaches revealed the impact of different molecular mechanisms on transcriptional bursting characteristics under specific experimental conditions. Using the approach of integrating a constant reporter at different genomic locations, several studies have shown that local chromatin environment affects mainly the burst size (Batenchuk et al., 2011; Singh et al., 2010; Skupsky et al., 2010). However, other studies taking wider observed expression ranges provided evidence of the impact of integration sites on both burst size and frequency in mammalian cells (Dar et al., 2012; Dey et al., 2015). Interestingly, Dar *et al*. found that enhancing expression with the cell signalling molecule TNF-α causes transcriptional burst frequency variation at low mRNA expression levels until a threshold is reached beyond which only burst size can then be further regulated (Dar et al., 2012; Nicolas et al., 2017).

Histone modifications appear to play a role in modulations of transcriptional kinetics. Analysis of transcription dynamics upon treatment with Trichostatin A (TSA, a histone deacetylase

inhibitor) revealed an increase in transcription rate with higher burst sizes observed (Harper et al., 2011; Suter et al., 2011). High acetylation levels, in other studies, increased burst frequency by shortening the duration of which the gene is inactive (Nicolas et al., 2017; Viñuelas et al., 2013). In addition, *cis*-regulatory DNA elements, such as TATA box and enhancers, help to a great extent shaping the bursting characteristics (Blake et al., 2006; Raser & O'Shea, 2004; Suter et al., 2011). The effects of different numbers and affinities of CCAAT box for the transcriptional activator NF-Y were examined in different cell lines; and results showed an increase in the burst size when using more than one CCAAT box or with higher affinity, without affecting the "on" phase duration (Suter et al., 2011). Studies looking at the role of enhancers using completely different approaches on different genes produced compatible results stating that enhancers control mainly burst frequency modulation rather than burst size (Bartman et al., 2016; Fukaya et al., 2016). Notably, recent work by Larsson *et al.* demonstrated, utilising scRNA-seq data, that burst size is encoded within core promoters, whereas burst frequency regulation is controlled via enhancer elements defining cell-type-specific variability (Larsson et al., 2019)

1.2 Gene expression variability in the innate immune system

Upon infection with pathogens, the mammalian innate immune system is required to produce a robust and effective response in order to protect the host without causing any self-damage (Chaplin, 2010). Therefore, tight regulation of immune responses is expected. However, cell-to-cell variability in innate immune responses is a feature that has been observed in different studies (Avraham et al., 2015; Rand et al., 2012; Shalek et al., 2013a; Zhao et al., 2012). Remarkably, Hu *et al.* work suggested that this variability is even essential for an effective innate immune response to viruses (Hu et al., 2011). However, understanding its functional importance in a broader context, how it is emerged despite the tight regulation of the system, and the mechanisms controlling it is still unclear.

The work presented in this thesis utilises the Toll-like receptor (TLR) signalling system, a well characterised evolutionarily-conserved class of pattern recognition receptors (PRRs) (Medzhitov, 2007), to study cell-cell variability in innate immune responses. For this reason, I

present an overview of the innate immune system, with an emphasis on the TLR, and what is known about cell-cell variability within this intricate system.

1.2.1 An overview of the innate immune system

The immune system is built upon two fundamental "lines of defence": the innate immune system and the adaptive immune system, covering different types of cells and complex processes. Although the two pillars operate through fundamentally different mechanisms, the inevitable interplay between them is vital for a robust, effective immune response (Chaplin, 2010). For the purpose of this thesis, I focus on the mechanisms of the innate immune system.

Innate immunity provides immediate, nonspecific defence mechanism against pathogens by employing its defensive barriers; physical (including skin and mucous membrane), chemical (such as lysozyme enzymes that break down bacterial cell walls), cellular (the focus of this thesis) which includes phagocytes and lymphocytes (Turvey & Broide, 2010). Phagocytes such as neutrophils and macrophages engulf and destroy pathogens through phagocytosis, while lymphocytes such as Natural killer (NK) cells can recognise and kill infected or abnormal cells. In turn, they play a role in activating the inflammatory barrier, the fourth defensive barrier, by producing key inflammatory proteins such as cytokines and chemokines (Marshall et al., 2018).

How does the innate immune system recognise pathogens?

Phagocytes and lymphocytes express PRRs on their surfaces or within their cytoplasm that can detect specific pathogen-associated molecular patterns (PAMPs) found on the surface of pathogens (Medzhitov, 2007), but not the host, allowing the immune cells to differentiate between self and non-self (Chaplin, 2010). A major class of PRRs is TLRs expressed on various cells, either on their cell surface or within endosomes. Macrophages are an example of TLR-expressing cells, they are key players in the innate immune system due to their ability to secrete cytokines that promote inflammation by recruiting other immune cells to the site of infection and activating their immune functions, as well as producing anti-inflammatory cytokines that regulate the inflammation response (Arango Duque & Descoteaux, 2014). Each TLR recognizes a specific PAMP (Akira et al., 2006). For example, TLR3 recognizes viral double-stranded RNA (dsRNA), a type of nucleic acid structure for several viruses, as well as its

synthetic version Polyinosinic-polycytidylic acid (PIC), also known as poly(I:C) (CHEN et al., 2021), while TLR4 is responsible for detecting bacterial lipopolysaccharides (LPS), a major component of the outer membrane of Gram-negative bacteria (Poltorak et al., 1998). Both TLRs are utilised in the work presented in the coming chapters.

Upon binding to their specific PAMPs, TLR3 and TLR4 recruit adaptor proteins such as Toll/interleukin-1 receptor (TIR)-domain-containing adaptor protein inducing interferon-β (TRIF) and myeloid differentiation primary response gene 88 (MyD88) (Bryant et al., 2015), which subsequently activate signalling pathways like Nuclear factor-kappa B (NF-κB) and the interferon regulatory factor (IRF) pathways, resulting in the transcriptional upregulation of various genes involved in immune responses (Bryant et al., 2015), including pro-inflammatory cytokines, such as interleukins (IL-1α, IL-1β, IL-6, IL-10, and IL-12), TNF-α and type I interferons (including interferon-beta (IFN-β), chemokines, like CCL2, CCL5, CXCL8, and CXCL10, and other immune mediators (Hayden et al., 2006; Kramer, 2016; Liu et al., 2017) (see **Figure 1.2** for an overview of TLR signalling pathway). These molecules promote inflammation, recruit immune cells, enhance defences, and coordinate the overall immune response to viral and bacterial

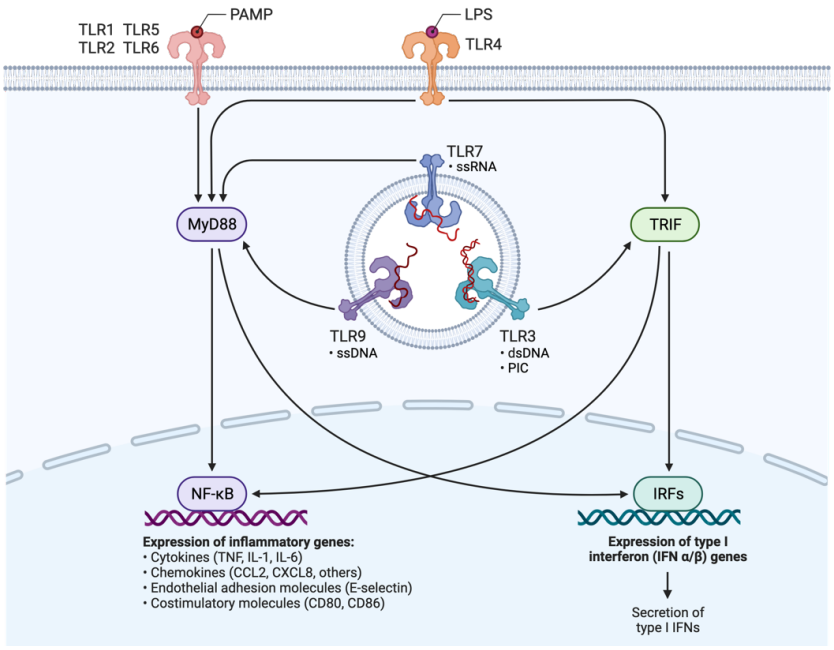


Figure 1.2: TLR signaling pathway. *Adapted from "TLR Signaling Pathway", by BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates. Shown are the different types of TLRs and their specific pathogen-associated molecular patterns (PAMPs). Upon their binding, adaptor proteins (such as MyD88 and TRIF) are recruited to activate signalling pathways (like NF-κB and IRFs) that are responsible for expressing immune genes.*

infection. In fact, TNF and IFN-I also have the ability to alter TLR4 responses through reprogramming the macrophages epigenome (Park et al., 2017).

Therefore, it is vital for the mechanisms responsible for producing these molecules to be tightly regulated, allowing no margin for error. But how can tightly regulated dynamic mechanisms result in heterogeneous immune response outcomes? This question serves as the central inquiry of this thesis.

Although the above-mentioned description of the innate immune responses is a very simplified version of all the mechanisms involved in the process, it still gives us a sense of the very complex network of biochemical events, including feedback loops, that underly precise immune outcomes.

1.2.2 Cell-to-cell variability in innate immune responses

Several studies demonstrated that, at population level, the TLR-mediated gene response is highly constrained related to tight epigenetic and transcription regulation (Adamik et al., 2013; Escoubet-Lozach et al., 2011; Hao & Baltimore, 2009; Meissner et al., 2013; Oda & Kitano, 2006a; Ramirez-Carrozzi et al., 2009; Tong et al., 2016). At the single-cell level, however, innate immune gene expression responses upon encounter with pathogens exhibit high variability (Avraham et al., 2015; Lu et al., 2015; Shalek et al., 2014; Xue et al., 2015). This is believed to reflect complex transcriptional regulation, involving dynamic TF signalling (Bagnall et al., 2018; Selimkhanov et al., 2014; Sung et al., 2014) as well as diverse genomic architecture (Hagai et al., 2018) in the host, in addition to cell-to-cell variability in the pathogen (Avraham et al., 2015).

In particular, Shalek et al., revealed, using scRNA-seq, bimodal expression patterns of many TLR-dependent genes across single cells upon stimulation with LPS (Shalek et al., 2013b). They also found that distinct cellular developmental states of dendritic cells, specifically maturity states, and stochastic differences in activating antiviral regulatory circuits contribute to the variability reflected in the bimodal expression. In a further study, the same group showed that stimulated bone marrow derived dendritic cells exhibit distinct patterns of variability (digital and analogue patterns, also seen in single-cell NF-κB dynamics (Tay et al., 2010)) determined by the stimulus and time of stimulation (Shalek et al., 2014), illustrating the importance of considering the temporal aspect when studying variability in gene expression.

Interferon (IFN)-mediated paracrine signals, which alter the repertoire of TF activation, seem to have a major role in regulating (promoting and restraining) this variability of TLR responses through positive and negative feedback loops (Shalek et al., 2014; Wimmers et al., 2018). In fact, as mentioned earlier, IFN-I has the ability to alter the epigenome of macrophages by modifying the chromatin structure to make it accessible and permissive of gene expression even at low levels of stimulation, this in turn prevents the occurrence of TNF "cross-tolerance", a phenomenon where cells become less responsive to TNF signalling over time. As a result, TNF can control target genes of NF-κB that encode inflammatory molecules (Park et al., 2017). It would not be surprising if this interplay between IFN-I and TNF to reprogram the epigenome contributes to cell-to-cell variability in TLR4 responses. NF-κB responses have been shown to be heterogeneous upon stimulation with *TNF*α (Tay et al., 2010). Gene expression variability of *TNF*α itself has been observed in macrophages after stimulation with lipid A, the cytotoxic component of LPS, and this variability has been explained partly by the differences in the cell size (Bagnall et al., 2018).

Another relevant source of variability in macrophages responses is the variability of the pathogens. Avraham et al., (Avraham et al., 2015) showed that high levels of IFN response genes are expressed in around only third of the cells, while the rest showing low levels of expression, confirming bimodal distribution observed in other parts of the immune system, as mentioned earlier. They combine fluorescent reporters, single-cell microscopy, and scRNA-seq methods to reveal that variability in macrophage IFN expression is attributed to variability in the infecting *Salmonella* cells rather than inherent variations within the macrophage population. This study provided a new dimension to the origins of cell-to-cell variability in the host system and added a new perspective in understanding host-pathogen interactions and outcomes.

Through the use of RNA FISH, interleukin-4 (IL-4), an anti-inflammatory cytokine, has been showed to produce mRNA transcripts upon stimulation in only 60% of cells population, and a significant proportion of these expressing cells were monoallelic with a small number of cells expressing two alleles (L. Guo et al., 2005). This is in agreement with several studies demonstrating that IL-4 expression is controlled by allele-specific activation, which in turn depends on the antigen dose (Bix & Locksley, 1998; Rivière et al., 1998). Monoallelic expression is the dominant mechanism until a sufficient increase in the signalling level

(reflecting high levels of antigen dose) that would allow for biallelic expression. Guo et al., (L. Guo et al., 2005) also proved experimentally that heterogeneity in frequency of IL-4 expressing cells and from which allele are primarily determined by probabilistic processes. They also argue that this stochastic regulation has a functional advantage in controlling differentiation of certain immune cell types upon interaction with IL-4.

Interestingly, an analogous mechanism to the antigen-dose-dependent activation of IL-4 alleles was revealed in macrophages upon stimulation with LPS. Macrophages has been shown to possess the ability to distinguish between different levels of LPS concentrations and to respond proportionally (Sung et al., 2014), playing an important functional role in controlling NF-κB activation dynamics. However, in this case almost all cells were responsive, contrary to the behaviour of NF-κB activity upon stimulation with *TNF*α, where the activation is heterogeneous at the single cell level, and the sensitivity to low doses is manifested in fewer responsive cells (Tay et al., 2010).

1.3 Methodologies for studying cell-to-cell variability

Throughout the years and with the accumulating interest in single-cell biology, experimental methods, including smFISH, scRNA-seq in fixed cells, and mass spectrometry tagging in living cells, have been developed and widely used to quantify transcription in single cells when studying mechanisms that underly the dynamics of transcriptional bursting, revealing how individual cells can be different from each other (Avraham et al., 2015; Dar et al., 2016; Ezer et al., 2016; Kim & Marioni, 2013; Newman et al., 2006).

1.3.1 Single-molecule Fluorescence *in situ* Hybridization

The single-molecule FISH technique provides measurements of the counts of individual mRNA molecules in a single cell as well as the number of active transcription sites (TS) (Femino et al., 1998). Fluorescent DNA probes bind to individual RNA molecules with specific complementarity, then static snapshots of mRNA molecules and active TS, appearing as bright dots with different intensities, are collected at different time points using fluorescence microscopy (Raj et al., 2008). Absolute number of mRNAs and active TS can be then counted using analysis tools like FISH-quant (Mueller et al., 2013). Other available software packages also allow quantifying transcription and degradation rates from smFISH images (Bahar Halpern & Itzkovitz, 2016). Several studies utilised this technique to analyse heterogeneity in

gene expression and revealed properties of transcriptional bursting (Little et al., 2013; Sepúlveda et al., 2016; Zenklusen et al., 2008) (as mentioned in previous sections).

1.3.2 Single-cell RNA sequencing

scRNA-seq experiments offer the ability to analyse the distribution of gene expression levels in all the expressed genes across a population of cells, which can range from hundreds to millions of cells. This allows for the investigation of cell-to-cell variations in transcription level within the population. There are various published experimental protocols (>20 to date) available for conducting scRNA-seq experiments. The selection of the most suitable protocol primarily depends on the specific research question. In general, scRNA-seq involves several key steps, including cell capturing, reverse transcription, amplification, library generation, and sequencing (Sasagawa et al., 2013; Tang et al., 2009). One major challenge faced by experimentalists when carrying scRNA-seq is striking the right balance between sequencing depth and the number of cells analysed (Zhang et al., 2020). Determining the optimal sequencing depth is crucial to ensure sufficient coverage of gene expression within individual cells, while considering the practical constraints of cost and time.

Strategies for cell capturing include microtitre-plate-based, microfluidic-array-based, and microfluidic-droplet-based. Microtitre-plate-based methods, such as fluorescent activated cell sorting (FACS), work on isolating cells into individual wells of the plate without damaging the cells in the process. One of this method's advantages compared to other methods is the ability to assess whether two or more cells have been mistakenly isolated into one well, as well as discarding any isolated damaged cells. In addition, several efficient technology platforms are available to generate scRNA-seq libraries, such as the 10x Genomics systems, which can capture up to tens of thousands of cells. Subsequently, computational pipelines are employed to pre-process and clean the obtained data (Hwang et al., 2018).

scRNA-seq protocols often suffer from technical limitations and inherent noise causing dropout events in the data, where genes are not detected or appear as zero counts in certain cells despite their expression in reality (Svensson, 2020). This in turn can pose challenges for downstream analysis and interpretation. Recent protocols are providing ways to overcome

these challenges (Ran et al., 2020; Sarkar & Stephens, 2021) and quantify data in the number of mRNA transcripts, making it suitable for mathematical modelling (Larsson et al., 2019).

1.3.3 Live cell imaging

A complete understanding of cell-to-cell variability necessitates the studying of gene expression dynamics at a mechanistic level within individual cells, and live cell imaging is the most suitable experimental approach for achieving this goal (Sung & McNally, 2011). By directly visualizing gene expression in real time, live cell imaging provides valuable insights into the temporal and spatial aspects of cellular behaviour, compared to fixed-cell approaches, enabling a deeper understanding of the sources and consequences of cell-to-cell variability. Dynamic imaging of transcription is feasible through utilising RNA labelling methods in live cells such as stem-loop labelling and fluorescence protein tagging by the MS2 coat protein (MCP) system (Wu et al., 2012). The MS2-MCP system has been widely used to study real-time kinetics of transcriptional bursts (Chubb et al., 2006; Golding et al., 2005; Yunger et al., 2010).

A recent breakthrough in real-time RNA imaging involves the utilization of CRISPR-Cas13 systems (Yang et al., 2019), offering a promising alternative to overcome limitations associated with conventional methods. By employing Cas13 along with RNA guides (gRNA) that specifically bind to the target RNA of interest, this approach assures no genetic modifications in the targeted cells (Yang et al., 2019), unlike the MS2-MCP method. While improvements are in progress (H. Cao et al., 2022; Yang et al., 2022), the method still suffers from high signal-to-noise ratio in comparison to MS2-MCP and guidelines for designing efficient gRNAs are still lacking (Yang et al., 2019). Nevertheless, it has already been used to reveal important features related to gene expression (Y. Huang et al., 2023).

1.4 Mathematical modelling

Mathematical modelling integrated with appropriate experimental measurements is crucial to refine our understanding of transcriptional bursting and the resulting cell-to-cell variability. Here I describe mathematical models of gene expression used in the literature, their differences, and limitations as well as numerical methods for model solving, simulations and

model selection. For the purpose of this thesis, I focus on modelling the transcription process of gene expression and neglect translation.

1.4.1 Kinetic models of gene expression

Based on the Central Dogma of Molecular Biology (DNA to RNA to protein), one can think of the simplest rate-kinetic model to describe the dynamics of mRNA production in the system using the following deterministic ordinary differential equation (ODE):

$$\frac{dm(t)}{dt} = k_t - k_d m(t), \qquad (1.1)$$

where m is the number of mRNA, rather than its concentration, in a cell over time *t,* k_t and k_d are synthesis and degradation rates, respectively. Solving the ODE for the initial condition $m(0) = 0$ gives the solution:

$$m(t) = \frac{k_t}{k_d}\,(1 - e^{-k_d t}), \qquad (1.2)$$

which demonstrates that mRNA production will accumulate over time following an exponential function with a time constant $\frac{1}{k_d}$, reaching its equilibrium (steady state) at $\frac{k_t}{k_d}$.

A change in one of the kinetic parameters, e.g., due to stimulus, would lead to a new steady state for the system reached in an exponential manner with its time constant. From this model, one can establish the important relationship between the half-life of an mRNA and its degradation rate (C.-Y. A. Chen et al., 2008; Hargrove et al., 1991; Lugowski et al., 2018):

$$t_{\frac{1}{2}} = \frac{ln(2)}{k_d}. \qquad (1.3)$$

This model has been utilized in the earlier stages of mRNA studies (Greenberg, 1972; Rodgers et al., 1985). However, over the years and with much greater understanding of gene expression, this model is no longer deemed sufficient to describe the dynamics of the process due to the inherently probabilistic nature of chemical reactions and the relatively small number of biochemical molecules involved in gene expression which are ignored in this deterministic model.

1.4.2 Stochastic models of gene expression

A stochastic model is indispensable when studying gene expression. By considering the involved biochemical reactions as probabilistic events, a stochastic model accounts for the inherent variability and randomness in the process. This provides a more accurate representation of the experimental observations and a better comprehensive understanding

of the complex dynamics and emergent properties of gene expression, such as cell-to-cell variability.

1.4.2.1 The telegraph model

Most genes, in prokaryotes and eukaryotes, produce mRNA in a "bursty", intermittent, non-Poissonian manner (Golding et al., 2005; Raj et al., 2006; Raj & Oudenaarden, 2008; Sanchez & Golding, 2013; Suter et al., 2011). The simplest stochastic model that captures such behaviour is the so called 'telegraph' model, a two-state switching model in which the gene, specifically its promoter, randomly transitions from an inactive to an active state in which transcription occurs. Inactive-to-active transitions occur with an activation rate k_{on} resulting in short periods of transcriptional activity generating mRNA molecules at a transcription rate k_t until the gene is "switched off" and returned, with deactivation rate k_{off}, to its inactive state. The mRNA transcripts are degraded at a rate k_d (see **Figure 1.3**). This model, therefore, accounts for the main four biochemical reactions in the process (promoter activation, inactivation, mRNA synthesis and mRNA degradation).

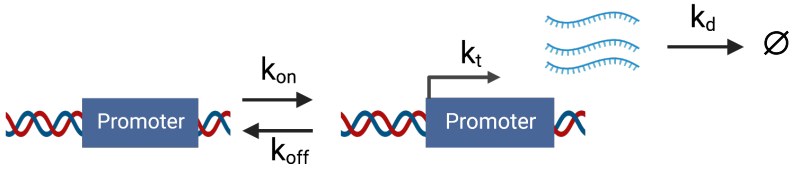


Figure 1.3: Schematic of the telegraph model. *Gene promoter can be in one of two states: inactive (left) or active (right). When in the active state, mRNA transcripts are produced at a rate k_t, which then degrade at a rate k_d. Transitions from the inactive to the active state occur at a rate k_{on}, while those from active to inactive go at rate k_{off}.*

Let $\tau_{on} = \frac{1}{k_{off}}$ be the time the promoter stays active for, while $\tau_{off} = \frac{1}{k_{on}}$ be the off-time. Then, under this model, burst size and frequency at steady state can be defined in terms of the model's kinetic parameters as follows (Nicolas et al., 2018):

$$b = k_t \tau_{on} = \frac{k_t}{k_{off}}, \qquad (1.4)$$

$$f = \frac{1}{\tau_{on} + \tau_{off}} = \frac{k_{on} k_{off}}{(k_{on} + k_{off})}. \qquad (1.5)$$

I refer to these as kinetic estimators and show in **Chapter 2** that they can be equivalent to the moment estimators ($\boldsymbol{b = \frac{\sigma^2}{\mu}}$, $\boldsymbol{f = \frac{\mu}{b-1}}$) under certain conditions.

In mathematical terms, the described model considers the temporal evolution of a gene state as a Markov process and is represented analytically using the Chemical Master Equation (CME). The CME is a set of coupled ODEs describing the time evolution of $P(x,t)$, the probability of the biochemical system being in a particular state x at time t; where a state x is defined by the number of mRNA molecules, i.e., probability of having m mRNA in a single cell at time t. Under the assumption that different promoters of the same gene behave independently (Skinner et al., 2016), and when experimental methods permit detection of an individual promoter's activity (Gómez-Schiavon et al., 2017), the number of active promoters (cells can have more than one active promoter) in a cell can also be incorporated in the CME such that the definition of the state becomes $\boldsymbol{x} = [n, m]^T$, where n is the number of active promoters in the cell.

The probability $P(\boldsymbol{x}, t)$ of being in each possible state then becomes:

$$\frac{dP(\boldsymbol{x},t)}{dt} = \sum_k P(\boldsymbol{x} - \boldsymbol{v}_k, t)\, a_k(\boldsymbol{x} - \boldsymbol{v}_k) - P(\boldsymbol{x}, t) \sum_k a_k(\boldsymbol{x}), \qquad (1.6)$$

where a_k is the propensity function such that $a_k \partial t$ is the probability that the biochemical reaction k will happen within the infinitesimal time interval ∂t depending on the rate of the reaction, and the stochiometric vector $\boldsymbol{v}_k$ describes the change in the system state, i.e., the change in the number of mRNA and promoters as a result of reaction k happening. It follows that $a_k(\boldsymbol{x} - \boldsymbol{v}_k)$ is the probability for the system to transition from state $\boldsymbol{x} - \boldsymbol{v}_k$ to state $\boldsymbol{x}$ through reaction k and the summation is over all the possible reactions. Detailed theory and a rigorous derivation of CME can be found in (Gillespie, 1992; Kampen, 2007; McQuarrie, 1967).

For the two-state telegraph model sketched above, the CME can be solved analytically using generating functions for when the system is in steady state (Raj et al., 2006; Shahrezaei & Swain, 2008b), as well as in the transient state (Iyer-Biswas et al., 2009). The steady state distribution has been shown to be the following (Peccoud & Ycart, 1995; Raj et al., 2006; Shahrezaei & Swain, 2008b):

$$P\left(m \middle| k_{on}, k_{off}, k_t, k_d\right) = \frac{\left(\frac{k_t}{k_d}\right)^m e^{-k_t/k_d}}{m!} \frac{\Gamma\left(\frac{k_{on}}{k_d}+m\right)\Gamma\left(\frac{k_{on}}{k_d}+\frac{k_{off}}{k_d}\right)}{\Gamma\left(\frac{k_{on}}{k_d}+\frac{k_{off}}{k_d}+m\right)\Gamma\left(\frac{k_{on}}{k_d}\right)} {}_1F_1\left[\frac{k_{off}}{k_d}, \frac{k_{on}}{k_d}+\frac{k_{off}}{k_d}+m; \frac{k_t}{k_d}\right], \quad (1.7)$$

where $\Gamma(.)$ is the gamma function and ${}_1F_1$ (*a, b; z*) is the confluent hypergeometric function.

While the steady-state mRNA first and second moments are given explicitly by the following expressions (Paszek, 2007):

$$\mu = \frac{k_{on}k_t}{k_d(k_{off}+k_{on})}, \tag{1.8}$$

$$\sigma^2 = \mu^2 \frac{k_{off}}{k_{on}\left(1+\frac{(k_{off}+k_{on})}{k_d}\right)} + \mu. \tag{1.9}$$

Having these moment expressions is useful, but not sufficient to analyse and gain a full picture of transcriptional bursting properties. A famous Monte Carlo algorithm known as (Gillespie's) stochastic simulation algorithm (SSA) is often used for numerically generating exact time trajectories for mRNA evolution in the system (see **Figure 1.4** for a realisation of SSA). The trajectories are based on the propensity functions and stoichiometry vectors of the possible reaction events, in which after every reaction the system state gets updated according to the stoichiometry of that specific reaction, with the assumption that the waiting times between events are distributed exponentially (Gillespie, 1977). The probability distributions of mRNA number can then be estimated from a sample of generated trajectories. It is noteworthy that although the SSA is easy to implement, the need to simulate every individual reaction event makes it very slow for large gene networks (more complex models). Therefore, improvements to the algorithm in order to speed it up have been established (Y. Cao et al., 2004; Gibson & Bruck, 2000; Lok & Brent, 2005; McCollum et al., 2006). In addition, the tau-leaping method (Y. Cao et al., 2006; Gillespie, 2007) is sometimes used as an approximate, but more computationally efficient, simulation strategy.

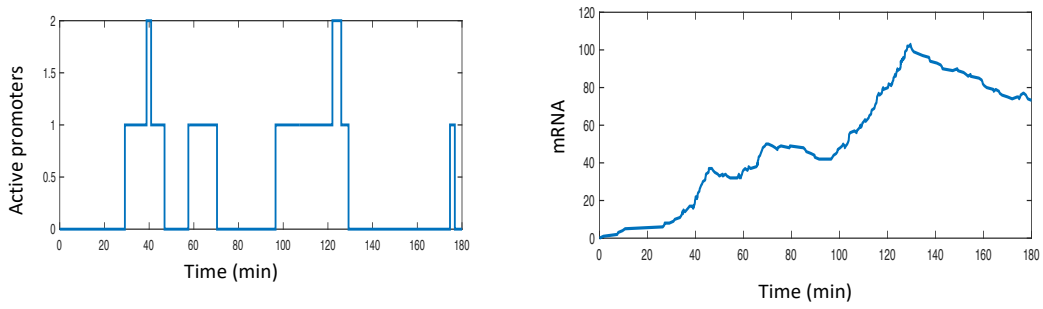


Figure 1.4: Stochastic simulation of gene expression using SSA. *Shown are trajectories of promoters' activity and mRNA temporal evolution according to the two-state model, simulated with k_{off} = 0.04 min^{-1}, k_{on} =0.02 min^{-1} and k_t=2 min^{-1}, k_d=0.01 min^{-1}.*

Is it possible to compute the exact solution of the CME?

In general, the CME can be written as:

$$\frac{dP(\boldsymbol{X},t)}{dt} = \boldsymbol{R}(\theta)P(\boldsymbol{X},t)$$

where $X = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_N}]^T$ is a vector of all possible cell states, N is the number of all possible states, $P(\boldsymbol{X},t) = [P(\boldsymbol{x_1},t), P(\boldsymbol{x_2},t), \dots P(\boldsymbol{x_N},t)]^T$ and $\boldsymbol{R}(\theta)$ is the state reaction matrix such that

$$R_{ij} = \begin{cases} -\sum_k a_k(\boldsymbol{x}_i) & \forall i = j \\ a_k(\boldsymbol{x}_j) & \forall j \text{ such that } \boldsymbol{x}_j = \boldsymbol{x}_i - \boldsymbol{\nu}_k \\ 0 & \text{otherwise.} \end{cases}$$

Time evolution of the probability distribution $P(\boldsymbol{X},t)$ is then given by: $P(\boldsymbol{X},t) = \exp[\boldsymbol{R}(\theta)t]\, P_0(\boldsymbol{X})$, where $P_0(\boldsymbol{X})$ is specified by initial conditions and $\sum_{\boldsymbol{X}} P_0(\boldsymbol{X}) = 1$.

When the number of states is small, computing the exponential of $\boldsymbol{R}(\theta)t$, and hence $P(\boldsymbol{X},t)$, is possible in MATLAB using a fast matrix exponential function (Al-Mohy & Higham, 2011). However, for a system with a large (up to 1000, which is possible for eukaryotes) number of mRNA transcripts, and therefore huge matrix $\boldsymbol{R}(\theta)$, computing $P(\boldsymbol{X},t)$ becomes time consuming. A clever method known as the finite state projection (FSP) algorithm comes to the rescue, providing an approximate solution by truncating the state space leaving out highly improbable states. A description of the algorithm is presented in (Senecal et al., 2014), including two examples based on a model of the Pap epigenetic switch demonstrating the application of the method in the biological field. Gómez-Schiavon *et al.* (Gómez-Schiavon et al., 2017) incorporated the FSP method in their work to calculate the post-stimulus distributional dynamics of the neuronal activity-inducible gene Npas4. Their approach can be generalized and used for calculating exact temporal mRNA distributions of other genes in which their expression can be recapitulated by the two-state model described above.

1.4.2.2 Other stochastic models of gene expression

Despite the common use of the telegraph model described in the previous section, it is apparent that this model is an oversimplification of the process and the complex biochemical events underlying it. It has yet to incorporate established processes which have been shown to affect gene expression, and therefore cell-cell variability, such as chromatin remodelling, methylation states of DNA, cell division and mRNA maturation. This led to several studies considering modified versions of the telegraph model by integrating some of the aforementioned biological processes. Senecal et al. (Senecal et al., 2014) extended the

standard two-state model by including a second ON state characterised with a higher initiation rate to the other ON state. They verified with experimental data on the early response gene *c-Fos* that this model accounts better for the distribution of nascent mRNA, and therefore they presented a model that describes the conversion of nascent mRNAs to mature mRNAs with a constant time delay. Mathematically, the probability being in the new ON state of the promoter over time is written as an ODE equation that is incorporated in the ODE system of two-state model.

In another study, Skinner et al. (Skinner et al., 2016) built upon this modified model to include the effect of cell-cycle and gene copy number differences between cells. They assigned a constant time period in which gene replication happens and during this time the gene can switch on with a specific rate, followed by dosage compensation at which this switching on rate, and therefore transcription, decreases by a calculated fold-change. They validated their model with experimental measurements on two mouse genes that control the pluripotency of embryonic stem cells, Oct4 and Nanog, and showed that Nanog has slower switching kinetics resulting in higher cell-cell variability. Through this model they provided a framework to study the effect of cell-cycle on cell-cell variability.

Although the above-mentioned models demonstrated an improvement over the telegraph model, their limitation lies in the fact that they can solely be analysed through stochastic simulations as their analytical solution remains unknown. Cao et al. (Z. Cao & Grima, 2020), on the other hand, presented a more comprehensive model that incorporates dosage compensation, replication, cell division, growth-dependent transcription and mRNA maturation, with the advantage of providing an analytical solution to the model.

Another model of gene expression incorporating features of promoter architecture is established by Zoller et al., (Zoller et al., 2015). They form a stochastic model that essentially looks like the two-state model but with N sequential inactive states reflecting refractory periods of silent transcriptional intervals (Cesbron et al., 2015; Harper et al., 2011; Suter et al., 2011).

1.4.3 Modelling gene expression data through statistical distributions

Another approach to model gene expression is through describing experimental gene expression data distributions using established probability distributions such as the lognormal (Bengtsson et al., 2005), Poisson (Sarkar & Stephens, 2021), negative binomial, gamma distributions (de Torrenté et al., 2020), or mixture models combining multiple probability distributions like the beta-Poisson model (a model of interest to us, as proved to capture bimodality featured in scRNA-seq data). The choice of the distribution depends on the nature of the data and the assumptions made about the underlying gene expression process. Nevertheless, this modelling approach provides a computationally efficient statistical framework for quantifying likelihood, spread, and shape of the expression levels across the population in various types of gene expression data, including bulk RNA-seq and scRNA-seq, providing a global view of gene expression patterns, while incorporating technical noise of the data (Anders & Huber, 2010; Love et al., 2014; Svensson, 2020).

Beta-Poisson mixture model

The formulation of the model is as follows: Let m be the observed count of mRNA molecules for a gene in a single cell. The beta-Poisson model assumes that m follows a Poisson distribution, $P(m) = \frac{e^{-\lambda}\lambda^{k}}{m!}$, where the mean parameter λ is itself drawn from a beta distribution $P(\lambda) = \frac{\lambda^{\alpha-1}(1-\lambda)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, with shape parameters $\alpha > 0$ and $\beta > 0$ (Dattani & Barahona, 2017; Smiley & Proulx, 2010).

In the context of the two-state model, the shape parameters of the beta-Poisson distribution can be interpreted as the rates at which the gene's promoter switches on and off, normalised by the mRNA degradation rate (I omit the degradation rate term in the denominator for the sake of brevity), and the mean of the Poisson distribution is scaled by the transcription rate. Specifically, $\lambda|k_{on}, k_{off} \sim Beta(k_{on}, k_{off})$ and $m|k_t, \lambda \sim Poisson(k_t.\lambda)$ (Vu et al., 2016). It follows from that the marginal distribution is equal to:

$$P(m; k_t\lambda) = P(m|k_t\lambda) * P(k_t) = \frac{(k_t\lambda)^m e^{-k_t\lambda}}{m!}\frac{\Gamma(k_{on}+k_{off})}{\Gamma(k_{on})\Gamma(k_{off})}\lambda^{k_{on}-1}(1-\lambda)^{k_{off}-1},$$

which yields to:

$$P(m) = \int_0^1 P(m; k_t\lambda)\, d\lambda = \frac{{k_t}^m}{m!}\ \frac{\Gamma(k_{on}+k_{off})}{\Gamma(k_{on})\Gamma(k_{off})}\int_0^1 \lambda^{k_{on}+m-1}(1-\lambda)^{k_{off}-1}e^{-k_t\lambda}\, d\lambda$$

$$= \frac{{k_t}^m}{m!} \frac{\Gamma(k_{on}+k_{off})\Gamma(k_{on}+m)}{\Gamma(k_{on})\Gamma(k_{on}+k_{off+m})} \, {}_1F_1\left[m + k_{on}, m + k_{on} + k_{off}, -k_t\right]. \quad (1.10)$$

Note that ${}_1\mathrm{F}_1(a, b, x) = \exp(x){}_1\mathrm{F}_1(b - a, b, -x)$ (Muller, 2001), which makes equation *(1.10)* equal to equation *(1.7).* Therefore, the beta-Poisson distribution is equivalent to the steady state distribution of the two-state stochastic model. Consequently, it is convenient to use the beta-Poisson model as a statistical framework when dealing with scRNA-seq data (Kim & Marioni, 2013; Larsson et al., 2019), as it is computationally less expensive than the stochastic model and able to reflect the bimodality in genes with slow switching rates, as well as long-tailed behaviour in the data distribution (Wills et al., 2013).

1.4.4 Parameter inference and model fitting

Determining the biophysical parameters that govern a biological process is crucial for understanding the underlying mechanisms driving the biological system. It also plays a pivotal role in enhancing the mathematical models employed to describe the process, resulting in more accurate predictions of system behaviour, and consequently facilitating better experimental designs.

Within our scope, to understand gene expression variability at single-cell level, it is important to study the role of kinetic parameters on fluctuations of number of mRNA transcripts. However, it is quite difficult, if not impossible in most cases, to measure these parameters directly through experiments. Therefore, parameter inference, estimating unknown parameters by fitting observed experimental data to mathematical models, is a prominent issue and widely discussed topic in the field (Ashyraliyev et al., 2009; Lillacci & Khammash, 2010). In mathematical terms, it is an 'inverse problem' that involves finding the most likely or optimal values for these parameters that best explain or fit the observed data. Although it can be challenging for complex models as multiple combinations of parameters can yield identical predictions, the estimation process is typically achieved through statistical methods, optimization algorithms, or a combination of both like the Expectation-Maximisation (EM) method. I discuss some of these methods with a focus on the most relevant ones to the work presented in this thesis.

1.4.4.1 Optimization algorithms

Different optimization methods such as genetic algorithms (Srinivas & Patnaik, 1994), least-squares fitting (Mendes & Kell, 1998), simulated annealing, and others (Ashyraliyev et al.,

2008; Kirkpatrick et al., 1983) are widely implemented when fitting model parameters. These methods search for the set of parameters that minimizes the distance (error) between the model distribution and the experimental data.

Genetic algorithms

As the name indicates, this inferring approach mimics the process of natural selection, the theory of biological evolution. Essentially it operates based on the principles of selection, reproduction, and mutation to iteratively improve a population of candidate solutions and find an optimal or near-optimal solution for a given problem (Srinivas & Patnaik, 1994).

In the scenario of inferring kinetic parameters of gene expression models, given a defined objective (fitness) function that measures the error between model distribution and experimental data, the genetic algorithm begins with an initial population of potential parameter sets. Each set represents a candidate solution. These parameter sets are then evaluated based on their fitness, using the objective function. The fittest parameter sets are selected for reproduction.

During reproduction, genetic operators such as crossover and mutation are applied to the selected parameter sets. Crossover involves combining genetic information from two or more parameter sets to create new offspring with a mix of characteristics. Mutation introduces small random changes to the parameter values to explore different regions of the solution space.

The process of selection, reproduction, and mutation are iteratively performed over multiple generations. With each iteration, the population evolves, and the fitness of the parameter sets improves, until converging to the optimal combination of parameter values that best fit the gene expression data (Ashyraliyev et al., 2009; *Genetic Algorithms*, 2021).

I use this algorithm in Chapter 2 and 3 to infer parameters of gene expression stochastic models. I minimise an objective function given by the average absolute distance between the theoretical (CME) and measured cumulative distribution functions (CDFs) across observed mRNA counts ($\frac{1}{n}\sum_{i=1}^{n}|CME_i - CDF_i|$), where i's are unique mRNA counts observed in the measured distributions. Several best model fits (50 for example in Chapter 2) are then considered to measure the Kaplan-Meier estimator of measured CDF and get a sense of the inference spread.

1.4.4.2 Bayesian inference approach

Despite the successful applications of optimization techniques in solving the problem of parameter estimation in several biological systems (Ezer et al., 2016; Kim & Marioni, 2013; Neuert et al., 2013), they can be computationally prohibitive and perform poorly if data includes significant noise (Lillacci & Khammash, 2010). Hence, state-of-the-art methods, such as Bayesian methods, are sought to infer kinetic parameters of gene expression models from single-cell data. Bayesian inference approach allows computation of the full probability distributions of the model parameters. The mathematical framework of this approach is based on the famous Bayes theorem, where given a stochastic gene expression model that depends on a set of unknown parameters θ and some observed data set D;

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \qquad (1.11)$$

where $P(D|\theta)$ is the likelihood of having sampled (or simulated if working with synthetic data) the observed data given the calculated time-dependent mRNA distributions for model parameters θ, $P(\theta)$ is the prior probability which is an expression of any prior knowledge and beliefs about parameters before observing any data, $P(D)$ is the probability distribution of the observed data which we call the evidence, and $P(\theta|D)$ is the posterior distribution: the probability distribution of the parameters given the observed data. In practice, computing $P(D)$ is often infeasible and Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis et al., 1953) and Gibbs sampler (Geman & Geman, 1984), are used to estimate the posterior distribution by sampling from the target distribution $P(D|\theta)P(\theta)$, the un-normalised version of the posterior (Suter et al., 2011; Zoller et al., 2015). Using the Bayesian approach, Gómez-Schiavon *et al.* (Gómez-Schiavon et al., 2017) established a novel computational pipeline (BayFISH) that allows inference of kinetic parameters of any stochastic gene expression model from smFISH data with quantification of their uncertainty given the data. In following work (Lin & Buchler, 2019), they developed and implemented an efficient hybrid Monte Carlo algorithm into BayFISH for speeding up the estimation of the posterior distribution.

While a Bayesian inference approach is the most complete method for parameter estimation, as it allows estimation of the entire probability distributions of the parameters instead of only

a point estimate, it still suffers from being computationally expensive for complex models, including multistate gene expression models.

1.4.4.3 Maximum likelihood estimation method

The underlying principle of the maximum likelihood estimation (MLE) approach is to select the set of parameter values that make the observed data most probable under the assumed statistical model, i.e., maximizing the likelihood function of the observed data (Rossi, 2018). The essential part of the method is defining the likelihood function.

Given a random sample (a set of observed data) $Y_1, Y_2, \dots, Y_n$ with density function f and under assumption of their independence, their joint probability distribution that is governed by a set of unknown parameters θ, $L(\theta)$ is the following:

$$L(\theta) = P(Y|\,\theta) = P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{i=1}^{n} f(y_i; \theta). \quad (1.12)$$

Therefore, the likelihood function depends solely on the type of data observed and the considered model of the process (Girolami, n.d.). For example, and in relation to the work of the upcoming chapters in this thesis, the likelihood function of single-cell data (smFISH and scRNA-seq) at several time points, given the model parameters, is a product of multinomial distributions, and is given by (Gómez-Schiavon et al., 2017):

$$P(Y|\theta) = \prod_{t=0}^{S} \left[\left(\frac{\left(\sum_j Y_j^t\right)!}{\prod_k Y_k^t!} \right) \prod_{i=1}^{N} [P(\boldsymbol{x}_i, \tau_t)]^{Y_i^t} \right], \quad (1.13)$$

where S is the total number of the data time points, Y^t is a vector of the number of cells displaying each observed state for sample at time t (the sum of this vector is the total number of cells N), $\boldsymbol{x}$ is the state of the cell defined by number of mRNA (and if applicable number of active promoters), and $P(x_i, \tau_t)$, depending on the considered model, can be derived using methods described earlier.

For numerical stability, the natural logarithm of the likelihood function, the log-likelihood denoted by $l(\theta)$, is usually considered. Therefore, MLE aims to estimate the model parameters over the parameter space Θ by solving the problem $\hat{\theta} = \arg\max_{\theta \in \Theta} l(\theta)$ (Haynes, 2013; Myung, 2003). This can be approached explicitly by setting the derivative of $l(\theta)$ to zero and solve the resulting equation, or numerically using optimization methods such as the profile likelihood approach (Pawitan, n.d.) and EM algorithm (Girolami, n.d.).

Larson et al. utilised the profile likelihood approach based on the beta-Poisson model to infer transcriptome-wide bursting characteristics using scRNA-seq data (Larsson et al., 2019). The efficiency of this method, particularly when dealing with large sample sizes, is one of its biggest advantages compared to Bayesian inference.

1.4.4.4 Model selection – "All models are wrong, but some are useful"

A model is a simplification or approximation of reality, driven by the knowledge we have about the process we are trying to model. The computational and experimental limitations we face and the lack of knowledge in very complex biological phenomena call for ranking the available models of the considered phenomenon from most useful to most useless, and therefore selecting the model that best explain the phenomenon under the circumstances of the study. Different methods are available to carry out the model selection task (Lillacci and Khammash, 2010), including several information criteria like Akaike information Criterion (AIC), Bayesian Information Criterion (BIC) and the Deviance Information criterion (DIC) which have been used successfully within a Bayesian inference framework (Gómez-Schiavon et al., 2017). However, it has recently been demonstrated that the use of Bayesian evidence incorporated with an Importance Sampler of the Harmonic Mean Estimator (IS-HME) (Robert and Wraith, 2009) to compute a model evidence could overperform in model selection compared to AIC and BIC particularly when relatively small sample size of data (~100 cells per time point) is considered (Lin and Buchler, 2019).

I use the AIC method in **Chapter 3** to assess different gene expression model fits and perform model selection. Explicitly, the AIC is calculated as follows; if the number of samples, n, is large enough:

$$AIC = 2k - 2ln\,(L) \qquad (1.14)$$

If $\frac{n}{k} < 40$:

$$AIC = 2k - 2\,ln(L) + \frac{2k(k+1)}{n-k-1} \qquad (1.15)$$

where L is the likelihood of the observed data given model parameters, and k is the number of parameters. The likelihood is defined by equation *(1.13)* and calculated by incorporating equation (1.10) for a beta-Poisson model, while incorporating the solution of equation *(1.6)* for stochastic models.

1.5 Aims and objectives

Cell-to-cell variability in gene expression, specifically mRNA level, is an intriguing biological phenomenon seen in all living organisms. While mathematical modelling and computational approaches have been used to understand the gene expression process in the broader context, several studies explored the biological mechanisms affecting transcriptional variability in isolation using experimental approaches. How individual innate immune cells adjust to changes of immune conditions while controlling gene expression patterns remains an unresolved question.

This thesis offers a novel understanding of how cellular variability of the innate immune system is mechanistically regulated upon facing different immune conditions through combining state-of-the-art mathematical modelling with smFISH and scRNA-seq experimental approaches. By investigating how transcriptional bursting features (burst size and frequency) are modulated in response to stimuli, this work provides a foundation for a more comprehensive understanding of the molecular mechanisms underlying the regulation of transcription at the single-cell level.

The main aims are:

- To provide novel understanding of how transcriptional heterogeneity in the TLR system is controlled under changes in immune conditions through single-cell experiments, and to develop a mathematical model that captures this phenomenon.
- To mathematically describe the modulation of bursting characteristics in response to the immune conditions' changes, and validate our theory through analysis of experimental data, thereby enhancing our understanding of the underlying dynamic mechanisms governing transcriptional heterogeneity.
- To investigate whether the heterogeneity observed in the innate immune system is predominantly driven by stochastic processes or if it is predetermined through transcriptional memory.

Chapter 2 Characterisation of transcriptional variability in the TLR system

Our understanding of how transcriptional heterogeneity in the TLR system is controlled remains incomplete. The initial part of this thesis is focused on addressing this question and developing mathematical tools and models that capture and allow analysis of this phenomenon. In this chapter, I explore characteristics of transcriptional variability in some TLR-immune genes. Single-cell experimental data suggests overall trends between the mean and variance of mRNA counts across different immune conditions. I show that these trends follow robust linear regression models presenting a solid framework to describe how transcriptional variability is constrained. Further, I introduce different stochastic models of gene expression describing the expression distributions of two immune genes TNFα and *IL*1β showing distinct levels of variability. Modulation of transcriptional bursting characteristics (burst size and frequency) that underly the linear trends are also calculated under the telegraph model regulatory mode. Experimental verifications of these calculations are presented.

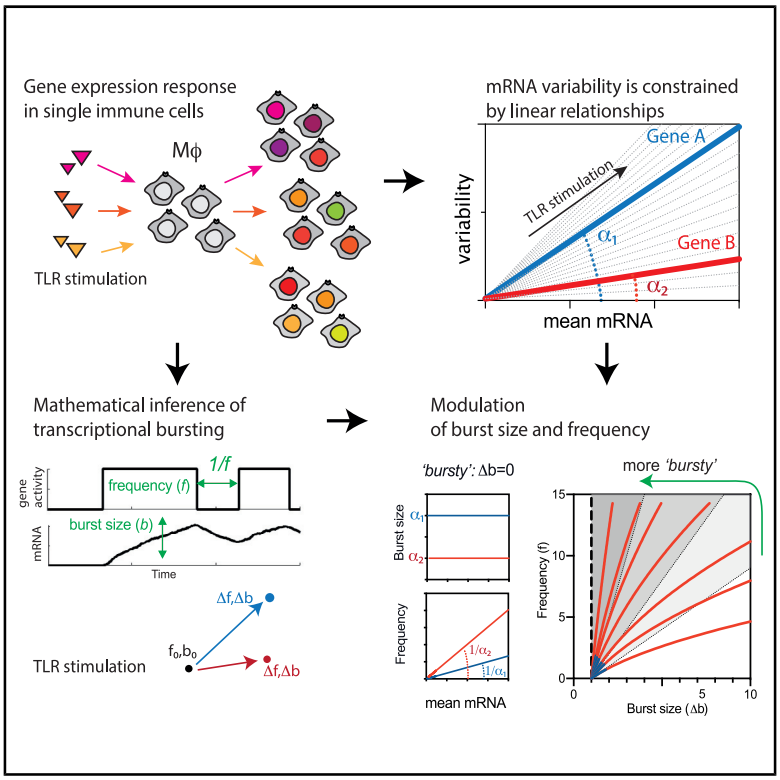
2.1 Journal paper: Gene-specific linear trends constrain transcriptional variability of the Toll-like receptor signaling

Cell Systems

Article

Gene-Specific Linear Trends Constrain Transcriptional Variability of the Toll-like Receptor Signaling

Graphical Abstract



Authors

James Bagnall, William Rowe, Nissrin Alachkar, ..., Dean A. Jackson, Mark Muldoon, Pawel Paszek

Correspondence

pawel.paszek@manchester.ac.uk

In Brief

This study demonstrates that transcriptional variability of the toll-like receptor signaling is constrained by mean mRNA responses through gene-specific modulation of transcriptional bursting.

Highlights

- Single-cell TNF-α and IL-1β mRNA responses are differentially controlled
- Variability of TLR-induced responses scale linearly with mean mRNA counts
- Gene-specific constraints emerge via modulation of transcriptional bursting
- Chromatin state regulates transcriptional bursting of IL-1β

Bagnall et al., 2020, Cell Systems *11*, 300–314
September 23, 2020
https://doi.org/10.1016/j.cels.2020.08.007

CellPress

CellPress
OPEN ACCESS

Cell Systems

Article

Gene-Specific Linear Trends Constrain Transcriptional Variability of the Toll-like Receptor Signaling

James Bagnall,[1,6] William Rowe,[2,3,6] Nissrin Alachkar,[1] James Roberts,[1] Hazel England,[1] Christopher Clark,[4] Mark Platt,[2] Dean A. Jackson,[1] Mark Muldoon,[5] and Pawel Paszek[1,7,*]

[1]Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester M13 9PT, UK
[2]Department of Chemistry, Centre for Analytical Science, Loughborough University, Loughborough LE11 3TU, UK
[3]Synbiochem, Manchester Institute of Biotechnology, University of Manchester, Princess Street, Manchester M1 7DN, UK
[4]Cancer Research UK Manchester Institute, University of Manchester, Wilmslow Road, Manchester M20 4BX, UK
[5]Department of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: pawel.paszek@manchester.ac.uk
https://doi.org/10.1016/j.cels.2020.08.007

SUMMARY

Single-cell gene expression is inherently variable, but how this variability is controlled in response to stimulation remains unclear. Here, we use single-cell RNA-seq and single-molecule mRNA counting (smFISH) to study inducible gene expression in the immune toll-like receptor system. We show that mRNA counts of tumor necrosis factor α conform to a standard stochastic switch model, while transcription of interleukin-1β involves an additional regulatory step resulting in increased heterogeneity. Despite different modes of regulation, systematic analysis of single-cell data for a range of genes demonstrates that the variability in transcript count is linearly constrained by the mean response over a range of conditions. Mathematical modeling of smFISH counts and experimental perturbation of chromatin state demonstrates that linear constraints emerge through modulation of transcriptional bursting along with gene-specific relationships. Overall, our analyses demonstrate that the variability of the inducible single-cell mRNA response is constrained by transcriptional bursting.

INTRODUCTION

Transcription of almost all mammalian genes is regulated by transitions in their association with active RNA polymerase complexes. This often results in brief periods of transcriptional activity and stochastic bursts of mRNA output characterized by their size and frequency (Raj et al., 2006; Raj and van Oudenaarden, 2008; Suter et al., 2011). Specific gene responses may exhibit different levels of heterogeneity, arising from variations in genome architecture (Dar et al., 2012; Dey et al., 2015; Nicolas et al., 2018; Zoller et al., 2015) in concert with regulatory signaling events (Larson et al., 2013; Megaridis et al., 2018; Wong et al., 2018), through "intrinsic noise" in the stochastic process as well as extrinsic differences between cells (Elowitz et al., 2002; Hilfinger and Paulsson, 2011; Sherman et al., 2015). A recent study (Larsson et al., 2019) demonstrated that while core promoter elements control burst sizes, regulation of bursting frequency via enhancer elements defines cell-type-specific expression variability. Similarly, histone acetylation can control burst frequency, but not burst size, to regulate the circadian gene output (Nicolas et al., 2018). It is generally assumed that single-cell, and thus, population-level responses to stimulation must be tightly controlled (Paszek et al., 2010; Stelling et al., 2004), although how this is achieved in the presence of the inherent noise is not fully understood. Analyses of gene expression from reporter cells suggest a paradigm where the noise of gene expression is inversely proportional to the mean expression level (Dar et al., 2012, 2016). However, these analyses rarely involve systematic perturbation of the same gene output and have not been performed on a genome-wide scale. Consequently, there is currently no clear understanding of how the variability of specific mRNAs change as a function of the magnitude of the response to acute stimulation or general perturbation.

In order to investigate the control of cellular variability, we used the well characterized toll-like receptor signaling (TLR) system (Medzhitov, 2007). TLR represents an acute innate defense mechanism against evolutionary-conserved pathogen-associated molecular patterns and involves a coordinated production of hundreds of genes, including pro-inflammatory cytokines

and chemokines (Bryant et al., 2015). The TLR effector response requires a fine balance between rapid yet robust immune activation while preventing out-of-control inflammation driving disease states (Bradley, 2008; Dinarello, 2011). Population-level studies suggest a highly constrained model, where the target gene response is subjected to a tight epigenetic and transcriptional regulation (Adamik et al., 2013; Escoubet-Lozach et al., 2011; Hao and Baltimore, 2009; Martin et al., 2020; Meissner et al., 2013; Oda and Kitano, 2006; Ramirez-Carrozzi et al., 2009; Tong et al., 2016). In contrast, at the single-cell level, TLR-dependent gene-expression responses exhibit high variability (Avraham et al., 2015; Lu et al., 2015; Shalek et al., 2013, 2014; Xue et al., 2015). This variability is thought to reflect complex transcriptional regulation, involving dynamic transcription factor (TF) signaling (Bagnall et al., 2018; Selimkhanov et al., 2014; Sung et al., 2014) as well as diverse genomic architecture (Hagai et al., 2018) and quorum licensing (Muldoon et al., 2020). For example, interferon (IFN) and tumor necrosis factor alpha (TNF-α)-mediated paracrine signals, which alter the repertoire of TF activation have been shown to regulate the heterogeneity of TLR responses (Shalek et al., 2014). However, the mechanisms by which the TLR system controls transcriptional bursting in order to regulate the heterogeneity of the target gene expression is not fully understood.

In this study, in order to uncover mechanisms that control gene-expression variability in the TLR system, we used single-molecule mRNA and single-cell RNA-seq (scRNA-seq) data obtained via systematic perturbation of individual gene outputs across immune-relevant conditions (Figure 1A). We specifically measured and mathematically modeled mRNA count distributions of TLR-dependent interleukin-1β (IL-1β) and TNF-α. We demonstrated that in response to 14 different TLR conditions the variability of the individual mRNA response can be empirically described by a linear function of the mean. These linear relationships are also present in 204 TLR-regulated genes in the scRNA-seq dataset from bone marrow dendritic cells (BMDCs) (Shalek et al., 2014). In the context of the stochastic telegraph model, we determined the ways in which the linear relationships constrain the underlying bursting characteristics. Theoretical predictions were subsequently validated by the analysis of TNF-α and IL-1β smFISH counts, including additional experimental perturbation of the chromatin state.

RESULTS

Expression of IL-1β and TNF-α mRNAs Exhibit Different Levels of Cellular Heterogeneity

To obtain insights into the control of cellular variability in the TLR system, we first characterized gene-expression patterns in innate immune macrophages by single-cell transcriptomics (Figure 1A). We generated single-cell RNA-seq libraries using the C1 Auto Prep System (Fluidigm C1) using an established RAW 264.7 macrophage cell line (Bagnall et al., 2018; Cheng et al., 2015; Sung et al., 2014) stimulated with lipid A for 3 h (the main cytotoxic component of TLR4 agonist lipopolysaccharides [LPS]; Raetz et al., 2007). After mapping and normalization (Figure S1), high-confidence genes (171 genes with higher expression and, hence, lower technical variance; Figure S1F), which were found to be regulated by lipid A in a previous population-level study (Bagnall et al., 2015), were clustered using an unsupervised affinity propagation method (Frey and Dueck, 2007). The analysis yielded 7 distinct major gene clusters and 3 uniform cell clusters (Figures 1B, S2A, and S2B; Table S1). For example, cluster XVII comprised 18 most abundant genes, including the effector cytokine *TNF-α* in addition to chemokines *Ccl9* and *Cxcl2*. Notably, we found a set of 10 genes that failed to cluster (referred herein as the "unclustered gene set"; Figure 1C). These included the pro-inflammatory inflammasome-associated cytokines *IL-1α* and *IL-1β* (Martinon et al., 2002) in addition to *IL-1rn* (interleukin-1 receptor antagonist), which are co-located in mouse and human genomes (Smith et al., 2004; Taylor et al., 2002). Other unclustered genes encoded chemokines: *Cxcl10*, *Ccl2,* and *Ccl5* and a pro-survival colony-stimulating factor *Csf3*, a ligand *Jag1* (Jagged1) (Hu et al., 2008), protein kinase (Plk2), a regulator of TNF-α secretion (Schwarz et al., 2014), and a membrane DC-stamp protein involved in cell fusion (Yagi et al., 2005).

Unclustered genes exhibited more variability than genes belonging to major clusters, while housekeeping genes were the most homogeneous (Figure S2C). Higher variation was not solely associated with technical noise as some major cluster genes have a higher number of mapped reads than the housekeeping genes (for example, clusters XVII and VII; Figure S2D). Similarly, unclustered genes do not have appreciably lower numbers of mapped reads than other genes and, indeed, have more in many cases. Expression heterogeneity may be related to physical gene properties (Hagai et al., 2018; Larsson et al., 2019), for instance, levels of transcriptional bursting have been linked to the presence of TATA boxes within gene promoters (Zoller et al., 2015). Indeed, we observe that unclustered genes exhibit significant enrichment of TATA sites in the promoter regions as well as a strong association between the transcript synthesis rate and variation (Figure S3).

We used quantitative smFISH to validate and accurately quantify expression patterns of *TNF-α*, *IL-1α,* and *IL-1β* mRNA in single cells (Figures 1D, 1E, and S4A–S4D). The average expression of *IL-1β* (± standard deviation, SD) was 215 ± 230 mRNA molecules for count data combined across all replicates. 50% of RAW 264.7 cells expressed more than 100 *IL-1β* mRNA molecules (with some expressing up to 1,000 molecules), while 20% of cells expressed <10 mRNA molecules (see Figure 1D for the cumulative probability function and Figure S4B for a histogram of smFISH counts). *TNF-α*, a cytokine that plays fundamental but distinct roles during infection (Adamik et al., 2013; Falvo et al., 2010), exhibited a similar level of expression on an average (255 ±144 mRNA molecules), but 90% of cells expressed more than 100 mRNA molecules (evident of reduced variability). We confirmed that the heterogeneous *IL-1β* expression patterns were seen in primary bone-marrow-derived macrophages (BMDM) (Figures 1E, S4C, and S4D), with correlated protein expression (Figures S4E–S4H) as well as in LPS-stimulated dendritic cells (Shalek et al., 2014) (Figure S5). There was also a good agreement between smFISH counts and our scRNA-seq study displaying similar levels of noise (Figure 1F). Overall, these analyses demonstrate conserved variability in the TLR system across cell types and suggest that *IL-1β* and *TNF-α* expression may have different modes of regulation.

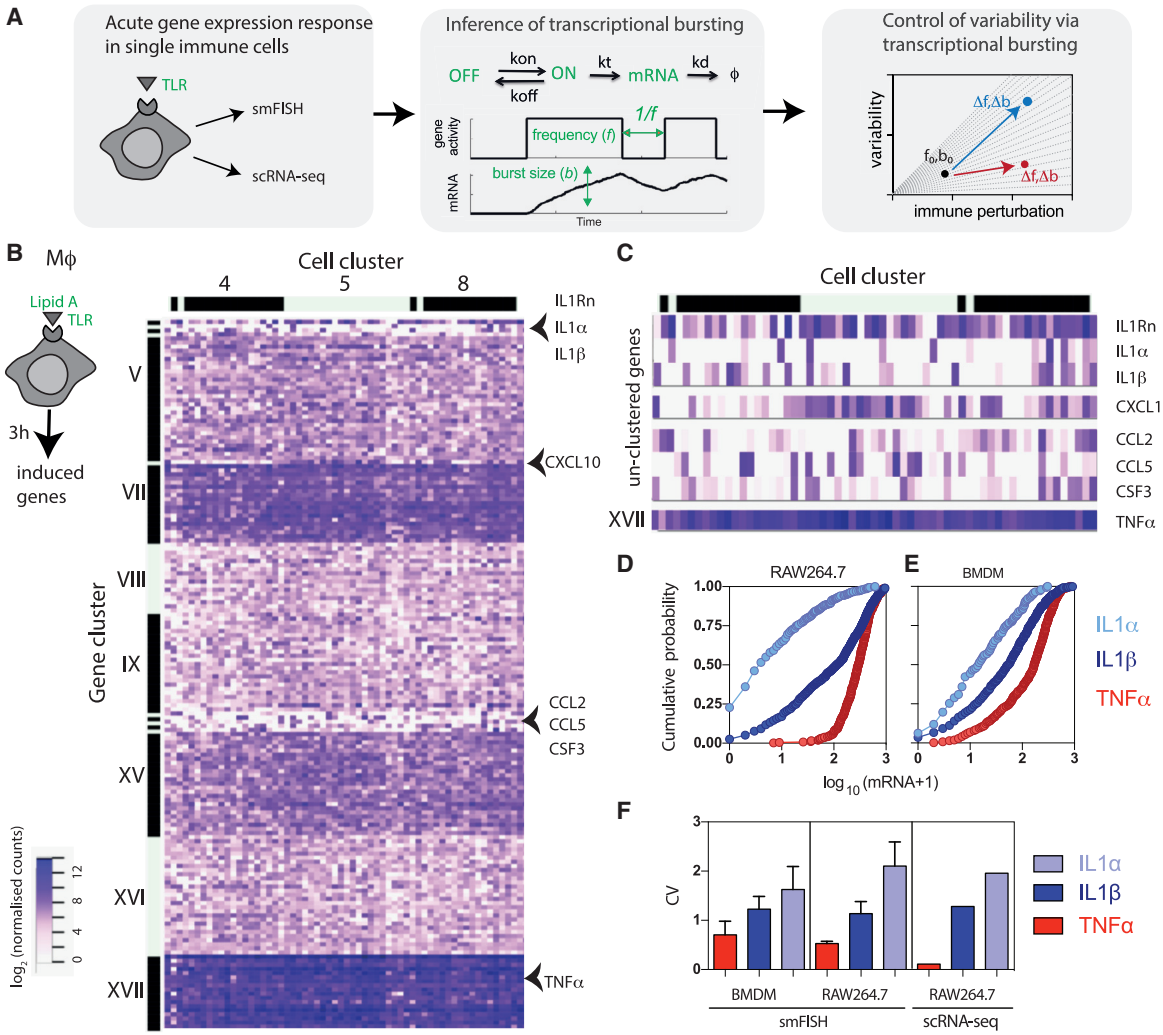


Figure 1. TLR4-Induced Effector Response Exhibit Differential Heterogeneity

(A) Schematic representation of the data analysis pipeline: gene-by-gene single-cell expression data are systematically analyzed across a range of immune-relevant conditions to understand the modulation of transcriptional bursting characteristics and control of cellular heterogeneity.

(B) scRNA-seq analysis of inducible TLR gene expression in RAW 264.7 cells stimulated with 500 ng/mL of lipid A for 3 h. Heatmap displaying normalized transcript levels of high confidence genes upregulated in response to lipid A stimulation. Major gene clusters are shown in roman numerals, cell clusters depicted with Arabic numerals. Arrowheads highlight specific unclustered genes as well as TNF-α.

(C) Heatmap of unclustered gene set from (B). Also shown is the heatmap of *TNF-α* expression.

(D) smFISH analysis of the cumulative probability distribution of *IL-1α*, *IL-1β*, and *TNF-α* mRNA expression in RAW 264.7 cells stimulated with 500 ng/mL of lipid A for 3 h. Count data expressed as $\log_{10}$(mRNA+1) from 447, 718, and 356 cells, pooled across at least three experimental replicates, respectively.

(E) Cumulative probability distribution of mRNA counts in BMDMs (stimulated as in D). Shown is the analysis of 447, 732, and 322 cells for *IL-1α*, *IL-1β*, and *TNF-α*, pooled across at least three experimental replicates, respectively.

(F) Variability of *IL-1α*, *IL-1β*, and *TNF-α* expression in scRNA-seq and smFISH data. Shown is the coefficient of variation (CV) calculated for respective genes across datasets, with SDs between biological replicates (when available).

Mathematical Modeling of mRNA Count Data Distinguishes Regulatory Modes

The heterogeneity of gene expression has typically been characterized in terms of transcriptional bursting, i.e., the process of intermittent gene activation (So et al., 2011). The characteristics of the transcriptional burst process, such as burst size and burst frequency, are defined as the average number of mRNA produced per gene activation event and the frequency of gene activation events, respectively (Nicolas et al., 2017). We first used the sample variance σ^2 and the mean μ of the mRNA distribution to compute an approximate burst size $b_m=\sigma^2/\mu$ (i.e., the Fano factor) and burst frequency $f_m=\mu/(b_m-1)$ (Nicolas et al., 2017; Raj et al., 2006; Suter et al., 2011) in order to understand the difference in *TNF-α* and *IL-1β* regulation. In general, these quantities (referred to here as "moment estimators") are often used to describe "burstiness" by quantitatively capturing departures from "non-bursty" (Poissonian) mRNA production (for which $b_m = 1$ *and* $f_m = \infty$) (Nicolas et al., 2017; So et al., 2011; Wong et al., 2018) (see Figure S6 for general applicability of the moment estimators). Analysis of the noise level ($CV = \sigma/\mu$), burst

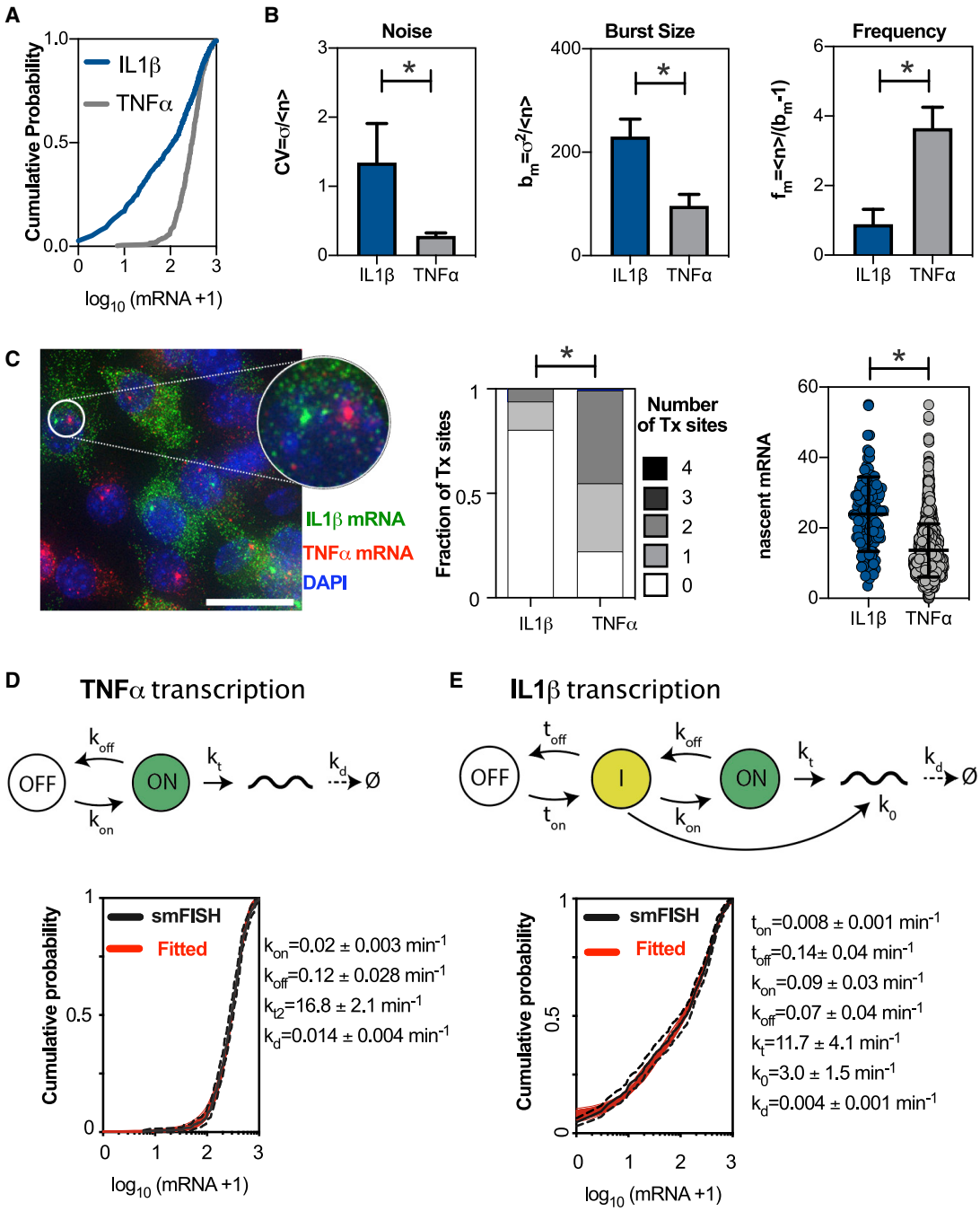


Figure 2. Mathematical Modeling Reveals Differential Control of *TNF-α* and *IL-1β* Transcription

(A) Differential expression of *IL-1β* and *TNF-α* mRNA. Shown is the cumulative distribution function of mRNA counts in RAW 264.7 macrophages stimulated with 500 ng/mL of lipid A for 3 h. A total of 718 cells were measured for *IL1β*, and 356 for *TNF-α*, and pooled across at least three smFISH experiments, respectively, and expressed as $\log_{10}$(mRNA+1).

(B) Characteristics of single-cell mRNA expression. Shown is the CV, burst size (b_m), and frequency (f_m) calculated based on moments of the mRNA count data from (A) (expressed as mean ± SD from experimental replicates). "*" denotes a result of a two-sample Mann-Whitney U test between groups ($p < 0.01$).

(C) Distribution of transcription sites is gene dependent. (Left) de-convolved wide-field microscopy image of cells with *TNF-α* and *IL-1β* smFISH, revealing Tx through an aggregation of multiple mRNA molecules in the nucleus (insert). Scale bar represents 5 μm. (Middle) the fraction of cells with 0–4 Tx calculated from (A). "*" denotes a result of the Fisher exact test ($p < 0.05$) for difference in the Tx site distributions. (Right) the number of nascent mRNA per Tx. Shown are individual Tx site data, together with the mean and SD of the pooled distribution. "*" denotes a result of a two-sample Mann-Whitney U test between groups ($p < 0.01$).

(legend continued on next page)

size (b_m), and burst frequency (f_m) based on the moments of the smFISH count distribution (Figure 2A) showed that *IL-1β* exhibits more burstiness, i.e., larger relative burst sizes and lower frequency compared with that of the more homogeneous *TNF-α* (Figure 2B). On-going *IL-1β* transcription, visualized via bright nuclear spots of fluorescence in the smFISH images (Femino et al., 1998; Skinner et al., 2016; Zenklusen et al., 2008), was evident in only 20% of cells (Figure 2C). In contrast, up to 75% of cells possessed at least one *TNF-α* transcription site (Tx). There was even an indication of *TNF-α* transcription immediately prior to cell division by the presence of >2 Tx sites in a subset of cells. We also observed more nascent mRNA associated with Tx sites for *IL-1β* than *TNF-α*. This is consistent with more infrequent but larger mRNA bursts in comparison to *TNF-α*. These characteristics were conserved across different doses of lipid A stimulation (including in BMDMs; Figures S7 and S8) as well as time (Figure S9), confirming that *IL-1β* and *TNF-α* exhibited distinct modes of transcriptional bursting.

The classical mathematical description of mRNA production involves a one-step stochastic telegraph model, where gene activity switches randomly between "off" and "on" states, with only the latter being permissive for mRNA transcription (Raj et al., 2006; Skinner et al., 2016; Suter et al., 2011; Zenklusen et al., 2008). The associated kinetic parameters include gene activation switching "on" and "off" rates (k_{on} and k_{off}, respectively) as well as rates of mRNA transcription and degradation (k_t and k_d, respectively); Figure 2D. In this case, bursting parameters are directly related to the kinetic parameters of transcription (Nicolas et al., 2018). The steady-state burst size is defined as $b_k = k_t/k_{off}$, while bursting frequency is given by $f_k = 2k_{on}k_{off}/(k_{on}+k_{off})/k_d$ (these are referred herein as "kinetic estimators"; Figure S6). In order to apply these stochastic models, we first investigated the sources of the variability in smFISH count data, which could either involve intrinsic stochastic fluctuations (i.e., on-off switching) or extrinsic cell-to-cell differences (Elowitz et al., 2002; Hilfinger and Paulsson, 2011; Sherman et al., 2015). We previously found a correlation between the cell size and mRNA level consistent with an extrinsic noise component (Bagnall et al., 2018), but this relationship did not affect mRNA distributions (when compared with cell size-normalized distributions) and only explained up to 7% of the data (as assessed by a correlation coefficient of a linear fit; Figure S10). Furthermore, smFISH counts exhibited a key intrinsic noise property, where noise decreased monotonically (Taniguchi et al., 2010) with mean expression, rather than approaching a plateau (Figure S11A). A formal noise decomposition of the *TNF-α* and *IL-1β* dose-response count data (Rhee et al., 2014) showed a dominant contribution from the intrinsic noise with an extrinsic noise component (Figure S11B). The latter is consistent with the extrinsic variability due to shared TLR signaling machinery, for example, signaling dynamics (Muldoon et al., 2020; Wong et al., 2018, 2019).

Given the dominant role of intrinsic noise, we, therefore, used a genetic algorithm to fit a family of one-step models (resulting in 50 kinetic parameter sets) to smFISH count distributions using biological constraints on parameter values (see STAR Methods). We found that the one-step model was able to recapitulate the measured *TNF-α* mRNA distribution in RAW 264.7 cells with an average gene switching "on" rate of $k_{on} = 0.02\ \text{min}^{-1}$ (i.e., equivalent to 50 min "off" time, $1/k_{on}$) and a switching "off" rate of $k_{off} = 0.12\ \text{min}^{-1}$ (i.e., equivalent to 8.3 min "on" time, $1/k_{off}$; Figure 2D). The average transcription rate of 16.8 ± 2 mRNA/min was consistent with the range previously reported for other highly inducible mammalian gene products (Molina et al., 2013; Schwanhäusser et al., 2011; Skinner et al., 2016; Suter et al., 2011) and was inversely correlated with the degradation rate (Figure S11C). We then used the one-step model to fit the distribution of *IL-1β* mRNA counts (Figure S11D), assuming a longer half-life in comparison to *TNF-α* (Hao and Baltimore, 2009). We found that the model failed to recapitulate the smFISH distribution, especially for mRNA counts below 100 molecules. We, therefore, considered more complex model structures that incorporate an additional constitutive initiation event, or additional regulatory step (equivalent to promoter cycling; Harper et al., 2011; Zoller et al., 2015), consistent with either chromatin remodeling or combinatorial TF binding driving a single transcription rate (Figure S11D models 2 and 3). These models were also unable to fit the observed data. Analysis of combined architectures suggested a model (Figures 2E and S11E) in which sequential activation and two transcription rates were required to recapitulate the entire range of mRNA counts. The first step was characterized by a small gene switching "on" rate $t_{on} = 0.008\ \text{min}^{-1}$ (equivalent of 125 min "off" time) and a low transcription output ($k_0 = 3 \pm 1.5$ mRNA/min); in contrast, the second step was rapid $k_{on} = 0.09\ \text{min}^{-1}$ (11 min "off" time) resulting in a high transcriptional output ($k_t = 11.7 \pm 4.1$ mRNA/min) (see Figure S11F for the comparison between individual on-off and transcription rates in the fitted family of models). During transcriptional activation, the first slow step is permissive for a second activation event resulting in a larger burst size and lower bursting frequency in the model, as compared with those for *TNF-α* (see Figure S11G for the estimates of the burst size and frequency from the models and Figure S11H for sensitivity analyses of model structures).

Transcriptional Heterogeneity Is Constrained by Gene-Specific Linear Trends

While our analyses demonstrate different levels of single-cell gene-expression heterogeneity in the TLR system, a fundamental question remains whether, and how, this heterogeneity is altered in response to stimulation or perturbation (Dar et al., 2012, 2016). In order to address this question, we systematically analyzed all smFISH datasets (Figure 3A) comprising the dose- and time-dependent responses in RAW 264.7 cells and BMDMs

(D) *TNF-α* transcription conforms to a one-step stochastic model. The comparison between measured and fitted *TNF-α* mRNA distributions at 3 h after 500 ng/mL lipid A treatment. In black: a Kaplan-Meier estimator of the measured cumulative distribution functions (CDF) (with 95% confidence intervals); and in red: a family of 50 models fitted to the data. Fitted parameter values (means ± SD) listed on the right.

(E) *IL-1β* transcription conforms to a two-step stochastic model. The comparison between measured and fitted *IL-1β* mRNA distributions at 3 h after 500 ng/mL lipid A treatment for the depicted model. In black: Kaplan-Meier estimator of measured CDF (with 95% confidence intervals); and in red: family of 50 models fitted to the data. Fitted parameter values (means ± SD) listed on the right.

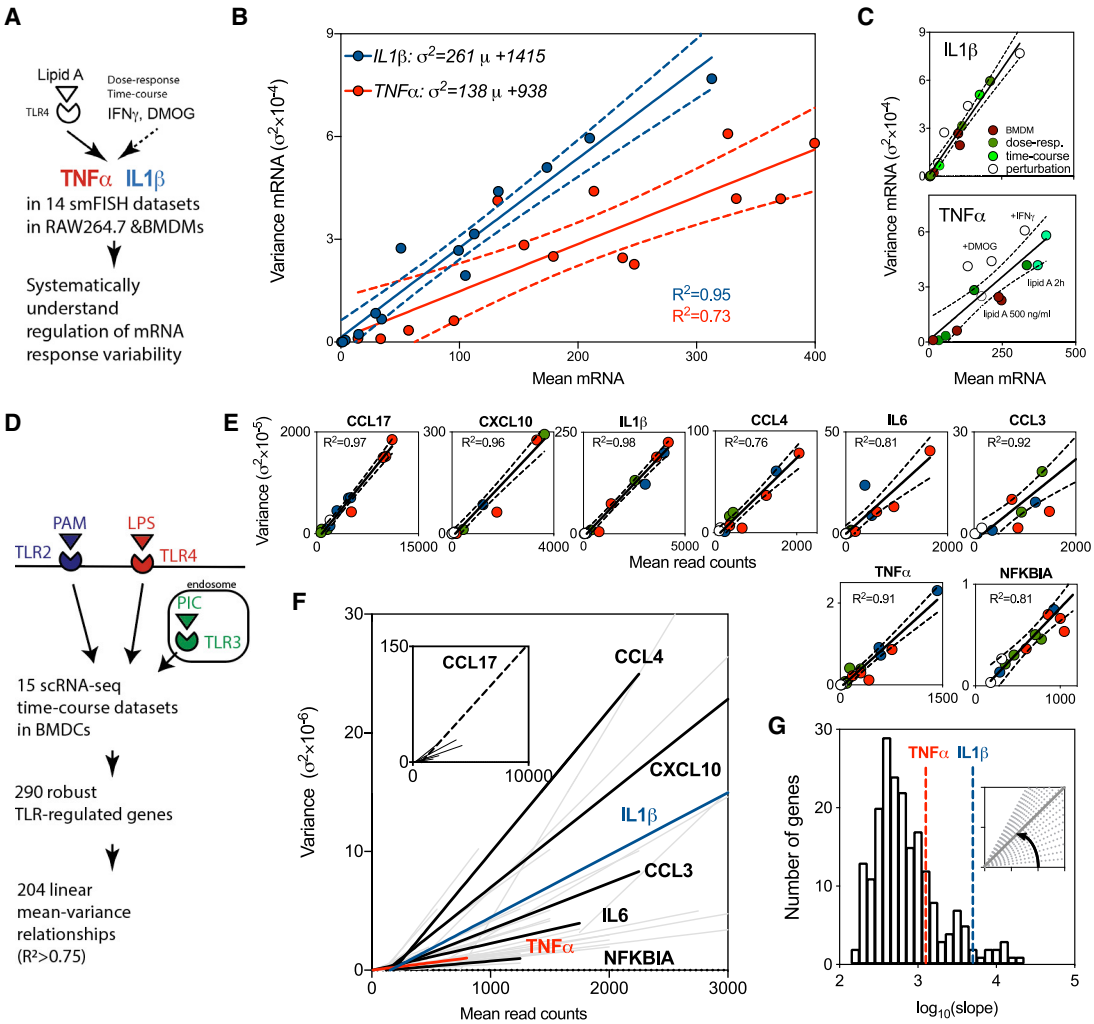


Figure 3. Single-Cell Expression is Constrained by Gene-Specific Linear Trends

(A) Analysis of single-cell variability in *TNF-α* and *IL-1β* mRNA expression across 14 smFISH measurements; dose response in RAW 264.7 and BMDM cells; time course in RAW 264.7 as well as DMOG and IFNγ co-stimulation in RAW 264.7 cells.

(B) Mean-variance relationship obtained for smFISH data for *IL-1β* and *TNF-α*. Shown is the fitted regression line (with 95% confidence intervals in broken lines), together with individual data points. Coefficient of determination depicted with R^2 and color coded. Fitted equations displayed on the graph.

(C) Visualization of samples across data in (B). Individual data points colored and labeled: green- RAW 264.7 dose-response data; light green, RAW 264.7 time course data; open circles, RAW 264.7, DMOG, and IFNγ co-stimulation data; and brown, BMDM dose-response.

(D) Inference of mean-variance relationships from the scRNA-seq data from (Shalek et al., 2014). BMDCs either untreated or stimulated with TLR2, 3, and 4 ligands for 1, 2, 4, or 6 h. For each TLR-dependent gene in the dataset, mean and variance of read count expression across all conditions are fitted using robust linear regression.

(E) Analysis of mean-variance relationships in selected TLR-induced genes. Shown are the fitted linear regression lines (with 95% confidence intervals) for highlighted genes from (Shalek et al., 2014). Different TLR treatments color coded as in (D) (open circles, untreated controls). Coefficient of determination depicted with R^2.

(F) Linear mean-variance regression trends for 204 high-confidence genes inferred from (Shalek et al., 2014). Highlighted genes depicted in black, trends for *IL-1β* and *TNF-α* in blue and red, respectively.

(G) Distribution of fitted regression slopes from (F) (in $\log_{10}$). Slopes for *IL-1β* and *TN-Fα* regression fits highlighted in blue and red lines, respectively.

to lipid A stimulation (Figures S7–S9) as well as additional immunologically relevant conditions (Figure S12). We used a 24-h interferon γ (IFNγ) pretreatment before lipid A stimulation, to mimic Signal Transducer and Activator of Transcription 1 (STAT1)-dependent inflammatory signaling (Bryant et al., 2015), which reduced *IL-1β* and increased *TNF-α* mRNA production (in comparison to stimulation with lipid A alone; Figures S12A and S12B). In turn, pretreatment with prolyl hydroxylase inhibitor dimethyloxalylglycine (DMOG), a pharmacological mimic of Hypoxia Inducible Factor 1α (HIF1α)-dependent hypoxia (Bagnall et al., 2014), resulted in an elevated expression of both *IL-1β* and *TNF-α* mRNA. When all smFISH datasets were examined collectively, we found that the gene-expression variability (represented as the variance of smFISH counts) across experimental conditions was constrained by the corresponding mean of the mRNA counts (Figure 3B). The larger heterogeneity in *IL-1β*

expression was reflected in a significant increase in the gradient of the mean-variance relationship (defined as a slope of the fitted regression line) than that of *TNF-α* (p value 0.00019). While some individual conditions showed departures from the fitted linear relationships (arguably more for *TNF-α* than *IL-1β*; Figure 3C), both fits were characterized by a high coefficient of determination (R^2 = *0.97* and *0.83*, for *IL-1β* and *TNF-α*, respectively). The fitted relationships appear to have positive intercepts, which is perhaps indicative of the limited sample size and might reflect measurement noise, therefore, we treat those as empirical relationships.

In order to establish this relationship in diverse cell types and in more genes, we took advantage of published single-cell transcriptomics data from BMDCs (Shalek et al., 2014); Figure 3D. These data included 15 scRNA-seq (each with up to 96 individual cells) time course measurements (at 0, 1, 2, 4, and 6 h) of acute responses to PAM (synthetic mimic of bacterial lipopeptides upstream of TLR2), PIC (viral-like double-stranded RNA for TLR3), and LPS (a component of Gram-negative bacteria upstream of TLR4), referred herein as the core TLR dataset. TLR pathways share common regulatory mechanisms, yet, induce distinct gene-expression patterns (Medzhitov, 2007). For example, the expression of *TNF-α* is maintained in response to PAM but is transient in response to PIC over the 6-h period (Figures S13A and S13B). However, in agreement with our smFISH data, we found that the mean and variance of *TNF-α* read counts exhibit a close linear relationship (Figure S13C; coefficient of determination R^2 = *0.91*). Subsequently, we considered 290 genes that were robustly induced by LPS stimulation in the dataset, revealing 204 genes that are described by linear trends with high confidence (as defined by R^2 >*0.75*; see Figure 3E for examples of specific genes and Figure 3F for the fitted relationships; Table S3 for all gene-by-gene fits). The previously observed trends in *IL-1β* and *TNF-α* expression were also present in the BMDC dataset (Figure 3F).

These analyses demonstrate that (1) the variability of mRNA expression can be empirically described by a linear function of the mean response; (2) the gene-specific variability can be defined by the slope of the regression line, constituting a spectrum at the genome level (Figure 3G). High variability genes include chemokines and cytokines, such as *CCL17*, *CCL3*, as well as *IL-1α* and *β*, while others, such as *TNF-α* (and *NFKBIA*, an inhibitor of NF-κB signaling) exhibit more homogeneous responses; (3) response patterns were shared among different TLR ligands and no difference between treatment-specific trends were found; (4) linear relationships were generally maintained under signaling perturbation involving Golgi inhibition and in Interferon-alpha receptr chain alpha (INFAR1), Tumour necrosis factor receptor 1 (TNFR), and STAT1 knockout cells (Shalek et al., 2014) (see Table S4 for gene-by-gene fits). However, the regression fit was altered in a subset of genes (as assessed by the analysis of regression slopes in the core TLR and perturbation datasets; Figures S13D–S13F), which suggests that these relationships can be regulated.

Linear Constraints Define Properties of Transcriptional Bursting

Previous studies suggest a paradigm where transcriptional bursting constrains stochastic gene-expression programs (Dar et al., 2016; Sanchez and Golding, 2013). The existence of an empirical linear relationship between the mean and variance of the single-cell mRNA response (Figure 3) provides insight into the regulation of transcriptional bursting. We used the steady-state approximation for the mRNA moments in the one-step model (Peccoud and Ycart, 1995; Paszek, 2007; Shahrezaei and Swain, 2008) and derived theoretical relationships between model parameters for which the $\sigma^2 = \alpha\mu$ relationship holds (Figures 4A, S14, and S15; STAR Methods for derivation and discussion). First, we considered the case of the "bursty" gene-expression regime, i.e., $k_{off} >> k_{on}$, when transcription occurs in short and infrequent bursts. Under these conditions, we theoretically predicted that bursting characteristics are predetermined by the empirical mean-variance relationship: (1) burst size is necessarily constant (and equal to the slope of the mean-variance line) over the range of the mean mRNA response (i.e., burst size $b_k = \alpha$-*1*); (2) changes of gene expression are controlled solely by frequency modulation [i.e., $f_k = \mu/(\alpha$-*1*$)$]; and (3) there is a reciprocal relationship between the burst size and frequency, as the burst frequency is proportional to the inverse of the burst size ($1/\alpha$). Therefore, the larger the burst size, the lower the frequency of gene expression (and vice versa) to maintain a constant mean-variance relationship. In a general case, our derivations show that both burst size and frequency may undergo modulation as the mean mRNA expression varies. The relative contribution of the burst size and frequency modulation is related to the k_{off} value (or k_{off}/k_{on} ratio; Figure S15). For a range of biologically plausible parameter values ($k_{off} < 0.2\ min^{-1}$ and $k_{on} < 0.1\ min^{-1}$, while $k_t < 30\ min^{-1}$), the higher the k_{off} (or k_{off}/k_{on} ratio), the smaller are the changes of the burst size in comparison to the changes of frequency (see Figure 4A for a set of putative genes with different levels of variability defined via slope α). For example, for $k_{off} > 0.1$ (and thus, relatively close to a bursty regime in the considered parameter ranges), we find 2-fold more changes of the burst frequency than that of the burst size (and 5-fold more for highly variable genes, i.e., $\alpha < 100$). In turn, $k_{off} < 0.02$ resulted in a dominant burst size modulation (especially for low variability genes).

In order to validate our theoretical predictions, we inferred bursting characteristics in our smFISH data across different immune-relevant conditions. First, using data from RAW 264.7 cells (to avoid cell-type differences), we fitted one-step models to measured *TNF-α* distributions. Consistent with the intrinsic noise model (Elowitz et al., 2002; Hilfinger and Paulsson, 2011; Sherman et al., 2015), *TNF-α* counts across all conditions fitted negative binomial distributions (Figure S16). Initially, we assumed a common half-life across all conditions (using k_d = *0.014* mRNA/min estimated for the high-dose 500 ng/mL lipid A treatment; Figure 2), while fitting three remaining parameters (k_{on}, k_{off}, and k_t; Figure S17). Later, all four kinetic parameters were refitted for a subset of conditions corresponding to lower lipid A doses (and thus, shorter mRNA half-life; Hao and Baltimore, 2009). Models summarized in terms of the mean-variance relationship (using fitted parameters to calculate moments; Figure 4B) captured most of the variability present in smFISH count data (with an exception of DMOG, which was subsequently not considered; Figure S17). Subsequently, we calculated the burst size and frequency changes along the fitted linear relationship (Figure 4C). Bursting characteristics were either obtained directly from fitted parameter values (using kinetic estimators) or predicted from the fitted regression line based on the fitted

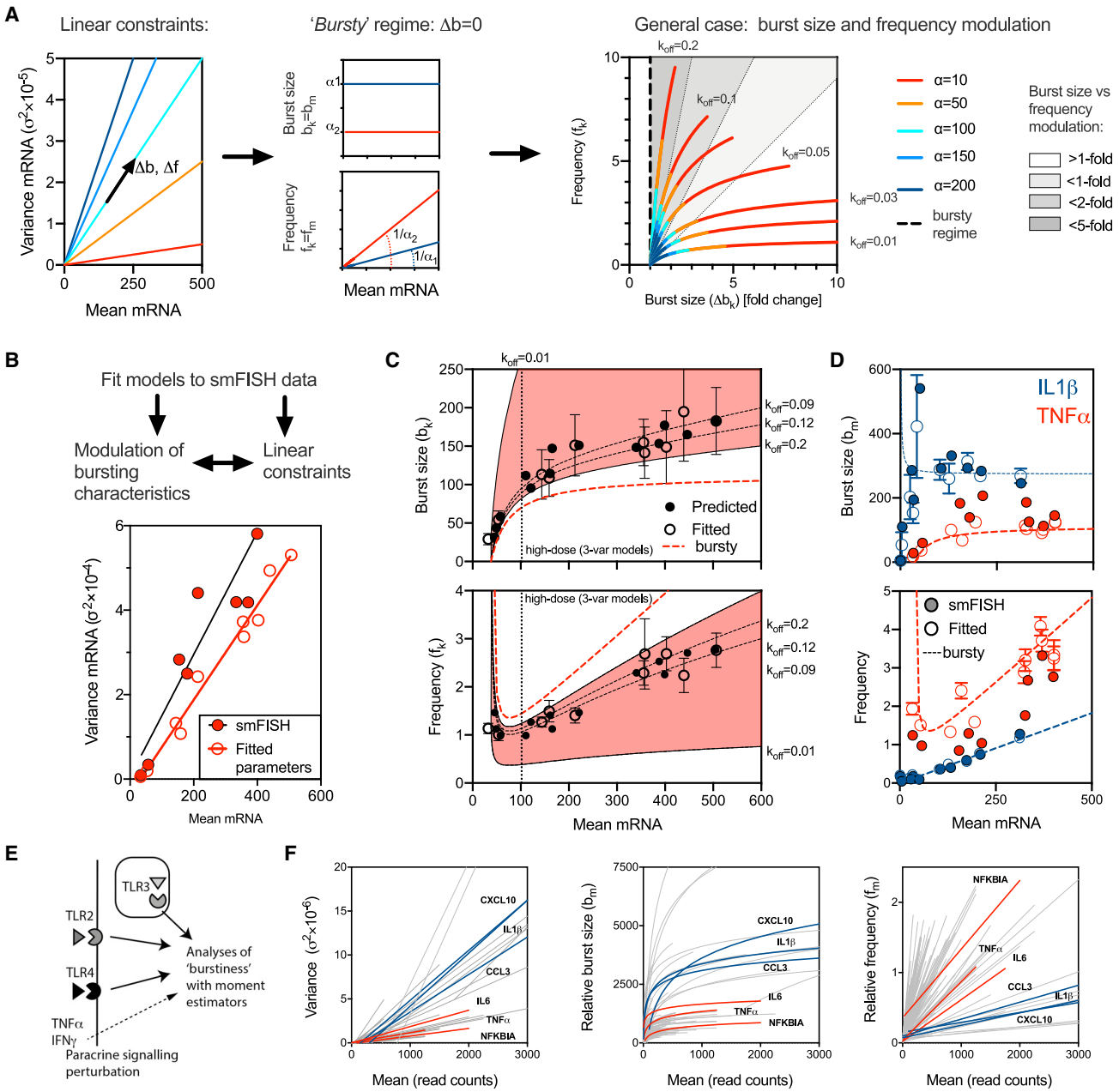


Figure 4. Linear Constraints Define Properties of Transcriptional Bursting

(A) Reciprocal relationship between burst size and frequency. (Left) a set of considered hypothetical genes characterized by different mean-variance slope α (such that $\sigma^2 = \alpha\mu$). (Middle) frequency modulation and constant burst size in the bursty regime. (Right) concurrent burst size and frequency modulation as a function of k_{off}. Calculations performed using Equation 6 for the biologically plausible set of gene activity switching rates, $k_{off} < 0.2\ min^{-1}$ and $k_{on} < 0.1\ min^{-1}$; $k_d = 0.014\ min^{-1}$; $k_t < 30\ min^{-1}$; and $\mu < 500$. Shown are relative frequency and burst size changes (Δb_k) over the corresponding range of the mean mRNA, calculated for each α for $k_{off} = 0.01, 0.02, 0.03, 0.05, 0.075, 0.1, 0.2\ min^{-1}$, respectively. In a broken line moment estimator (i.e., bursty regime), shaded are regions corresponding to 1-fold, 2-fold, and 5-fold burst sizes versus frequency modulation.

(B) Variability of the TNF-α expression across data in RAW 264.7 macrophages (dose response, time course, as well as IFNγ, IFNγ+lipid A, and DMOG+lipid A perturbation). Displayed is the relationship between sample mean and variance of individual smFISH count data (full red circles) and steady-state mean and variance (open red circles) based on fitted parameter values (Figure S17). Model outputs calculated for a family of 50 models fitted to each data point. Regression lines fitted to smFISH counts (depicted in black) and steady-state mean and variance calculated for fitted model parameters (depicted in red).

(C) Burst size and frequency modulation of the *TNF-α* expression. Shown in red are regions calculated for the fitted $\sigma^2 = 113\ \mu\text{-}4249$ relationship for biologically plausible set of gene activity switching rates: $k_{off} < 0.2\ min^{-1}$, and $k_{on} < 0.1\ min^{-1}$; and $k_d = 0.014\ min^{-1}$, and $k_t < 30\ min^{-1}$. Highlighted broken lines correspond to burst size and frequency changes corresponding to $k_{off} = 0.01, 0.09, 0.12, 0.2\ min^{-1}$. Predicted burst sizes and burst frequencies depicted in black circles (using Equation 7 and fitted k_{on}/k_{off} and k_d rates, from Figure S17), in open circles steady-state estimates using fitted parameter values. The broken red line shows a

(legend continued on next page)

k_{off} and k_{on} rates (see STAR Methods for derivation and discussion of a general case of mean-variance relationships with a non-zero intercept). Both approaches demonstrate a concurrent modulation of *TNF-α* expression via the burst size and frequency as a function of mean mRNA expression. The burst size increased monotonically from ~25 to ~150 molecules across all conditions, while the burst frequency changed between 1 and 3 (with a minimum predicted for the case of a linear fit with a non-zero intercept; Figure S15E). When considering only the subset of conditions for the high-dose lipid A responses (mean mRNA > 100 and $k_{\sigma} = 0.014$ mRNA/min), the changes of the burst size were limited to <2-fold. In this case, the frequency modulation becomes dominant in agreement with the theoretical prediction (bursty regime shown by the broken red line; Figure 4C). Analysis of fitted parameters demonstrates that the modulation of bursting characteristics across the mean expression was due to an increase in the "on" rate, and a concurrent decrease in the "off" rate (Figure S17C).

Our analyses predict a link between the level of expression variability and bursting characteristics, i.e., increased variability results in increased burst size and lower burst frequency (Figure 4A). Therefore, to compare gene-specific characteristics we fitted *IL-1β* smFISH count data in RAW 264.7 cells using the previously developed two-step model (Figure S18). Given the multistep structure of the *IL-1β* model, we reverted to moment estimators (Nicolas et al., 2017; So et al., 2011). In agreement with our modeling predictions, the higher variability of *IL-1β* expression is associated with quantitatively larger burst size and lower frequency (obtained via moment estimators) than that of *TNF-α* (see Figure 4D for relationships using fitted models and smFISH count data in RAW 264.7 cells- Figures S19A and S19B for analysis of all smFISH counts). In a general case, both burst size and frequency may undergo modulation, which is evident from the analysis of the *TNF-α* regulation (Figure 4C). Our analyses predict that the contribution of the burst size modulation decreases as the system converges to the bursty regime. *IL1β* transcription exhibits more "bursty" expression in comparison to *TNFα* (Figure 2, with t_{off}/t_{on} ~18 for *IL1β* in the permissive step and k_{off}/k_{on}~6 for *TNFα*). We find evidence for more a dominant frequency modulation of *IL1β* expression when compared with transcriptional bursting characteristics inferred for *TNFα* (at least for high mRNA expression; Figure 4D). In agreement, the burst size of *IL-1β* mRNA production remained constant for a wide range of expression (except for small means; see STAR Methods for discussion of mean-variance with a non-zero intercept). Consistently, the fitted parameter values exhibit changes in the gene activity switching "on" rates (corresponding to both regulatory steps) over the whole range of *IL-1β* mRNA responses (Figure S18D).

Finally, we used the scRNA-seq dataset in BMDCs (Shalek et al., 2014) to gain insights into the modulation of bursting characteristics in 323 robustly expressed TLR-dependent genes (Figures 4E, 4F, and S19B–S19F). We did not make any assumptions about the transcriptional regime (since fitting models to scRNA-seq dataset was not possible due to the lack of absolute quantification in the sequencing protocol; Shalek et al., 2014), but instead, we used regression analyses to infer changes of relative burst size and frequency (described by moment estimators) across gene-specific linear relationships (see Figure S19C for inference of bursting characteristics for *TNF-α*; Tables S3, S4, and S5 for gene-by-gene visualization, including a comparison between core TLR and perturbation datasets). Despite the inherent variability of the scRNA-seq data (which was validated by remapping a subset of data; Figure S20), we found quantitative changes of burstiness across >130 individual genes consistent with relative burst size and frequency modulation (Figure 4F). By comparison of independently fitted relationships, we also found that the evidence for the predicted reciprocal bursting characteristics, including a negative correlation between the burst size and frequency as well as the correlations with the slope of the fitted mean-variance relationships are present in the dataset (Figure S19F).

Chromatin Regulates *IL-1β* Expression via Modulation of Bursting Characteristics

Given the role of modulation of transcriptional bursting in the control of the inducible single-cell gene-expression variability, we sought to investigate the underlying mechanism. Previous work indicated the involvement of TF signaling, including that of the nuclear factor κB (NF-κB) in the TLR system in the context of chromatin regulation (Larson et al., 2013; Nicolas et al., 2018; Wong et al., 2018). We, therefore, turned our attention to *IL-1β* transcription, the two-step structure of which suggests an influence of the chromatin state. We observed a highly correlated biphasic mRNA response between *IL-1β* and *IL-1α*, whose genes are located in a single gene cluster in the mouse and human genome (Smith et al., 2004; Taylor et al., 2002), but not *TNF-α* (Figure S21A). We also observed a significant correlation between the presence of *IL-1β* and *IL1α* (but not *TNF-α*) active transcription sites (Figure S21B). We found that the transcription of *IL-1β* and *IL-1α* not only coincided temporally but also spatially, as a significant number of Tx sites co-localized, Figure S21C). Presumably, these genes sharing a local chromatin structure show a high propensity to be transcribed within a common transcription factory (Jackson, 2003).

The observed temporal and spatial coordination of *IL-1β* and *IL-1α* expression is suggestive of epigenetic mechanisms. A

predicted behavior in the bursty regime based on the fitted regression line. Horizontal dotted line marks a subset of data corresponding to the high-dose lipid A conditions (3-variable model fits; Figure S17).

(D) Burstiness of the *IL-1β* and *TNF-α* mRNA expression. Shown are moments estimates of burst size and frequency for smFISH counts (full circles) and fitted model distributions (open circles, in blue and red for *IL-1β* and *TNF-α*, respectively) for data in RAW 264.7 macrophages (dose response, time course, as well as TSA, IFNγ, and DMOG perturbation; Figures S17 and S18). In broken red and blue lines is the predicted behavior in the bursty regime, based on the regression lines for fitted models for *TNF-α* (from B) and *IL-1β* (from Figure S18C), respectively.

(E) Schematic representation of the combined (core TLR and paracrine signaling perturbation) scRNA-seq datasets from (Shalek et al., 2014).

(F) Burstiness of TLR-induced genes. Shown are relationships for the variance, relative burst size (b_m), and relative frequency (f_m) as function on the mean read count inferred from the combined core TLR and perturbation dataset from (Shalek et al., 2014). Displayed are 204, 180, and 132 relationships for variance, relative frequency, and relative burst size (defined based on the coefficient of determination $R^2 > 0.75$, $R^2 > 0.7$ and $R^2 > 0.5$, respectively) inferred using robust linear regression (with semi-log transformation for relative burst size). Individual high and low heterogeneity gene fits color coded and labeled.

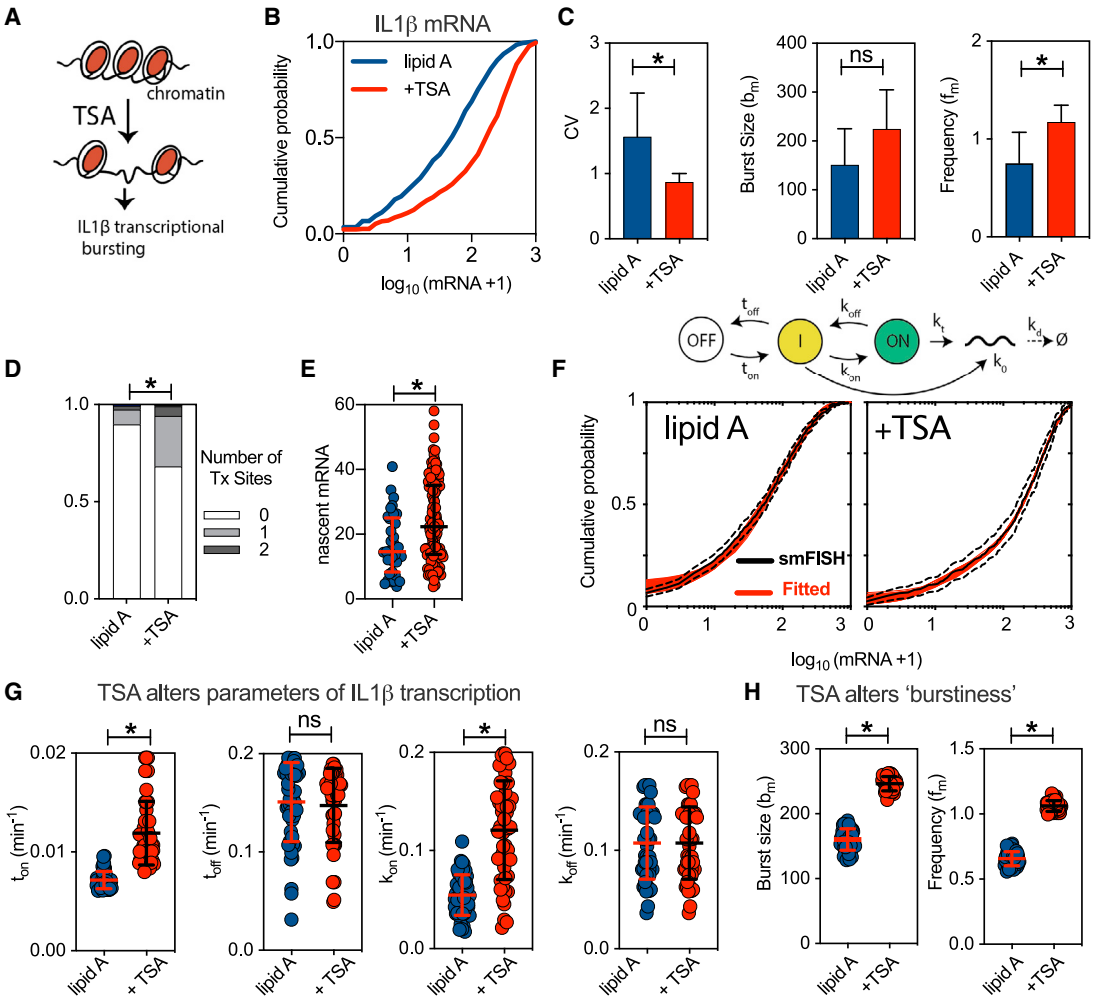


Figure 5. Modulation of Transcriptional Bursting via Chromatin State

(A) Schematic representation of the treatment protocol: cells exposed to 10 μM TSA for 1 h before 500 ng/mL lipid A treatment.

(B) TSA alters IL-1β mRNA distribution. Cumulative probability distribution of smFISH mRNA counts in BMDMs pre-treated with TSA prior to lipid A stimulation (+TSA; as in A), or control cells stimulated with lipid A. Shown is the *IL-1β* levels expressed as $\log_{10}$(mRNA+1) pooled across at least three replicates, from 732 (lipid A) and 305 (lipid A +TSA) cells, respectively.

(C) Characteristics of single-cell mRNA expression. Shown is the CV, b_m, and f_m calculated based on moments of the mRNA count data from (A) (expressed as mean ± SD from experimental replicates). "*" denotes a result of a two-sample Mann-Whitney U test between groups ($p < 0.05$; ns, not significant).

(D) Distribution of Tx in data from (B). Shown is the fraction of cells with 0–2 Tx. "*" denotes a result of the Fisher exact test ($p < 0.05$) for the difference in the Tx site distribution.

(E) Nascent IL-1β mRNA counts (with means and SDs) from 35 (lipid A) and 114 (lipid A +TSA) Tx from (D), respectively. "*" denotes a result of two-sample Mann-Whitney U test between groups ($p < 0.05$).

(F) Comparison between the measured and fitted *IL-1β* mRNA counts across conditions from (B). In black: Kaplan-Meier estimator of the measured CDF (with 95% confidence intervals); and in red: a family of models (50) fitted to the data. (Top) schematics of the fitted transcriptional model.

(G) TSA modulates kinetic parameter rates in the fitted *IL-1β* models. Shown are selected parameter values (with mean and SD) for families of fitted models from (F). "*" denotes a result of a two-sample Mann-Whitney U test between groups ($p < 0.0001$, ns).

(H) TSA alters bursting characteristics of *IL-1β* expression. Shown are the moment estimates (mean and SD) of the burst size and frequency for fitted mRNA distributions from F. "*" denotes a result of a two-sample Mann-Whitney U test between groups ($p < 0.0001$).

transcriptional activator trichostatin A (TSA) was applied to selectively inhibit the class I and II histone deacetylase (HDAC) enzymes responsible for genome-wide chromatin accessibility (Figure 5A) (Vanhaecke et al., 2004). BMDMs pre-treated with TSA for 1 h prior to 3 h lipid A stimulation exhibited elevated *IL-1β* expression, notably, the expression of *TNF-α* was completely abolished (Figure 5B). The resulting *IL-1β* mRNA distribution was shifted toward higher mRNA counts (in comparison to the lipid A control) (Figure S22 for the lipid A dose response). TSA pretreatment significantly reduced the noise of *IL-1β* expression and altered burstiness by significantly increasing the moment estimate of bursting frequency, there was also an indication of changes in the burst size (Figure 5C). The number of active Tx sites was increased, consistent with more frequent

activation of transcription following TSA treatment (Figure 5D). In comparison with the lipid A treatment, each *IL-1β* Tx site was also associated with more nascent mRNA (Figure 5E) indicative of a larger burst size. The differences in bursting characteristics were maintained across different lipid A doses (when co-treated with TSA), while the corresponding mean-variance relationships could not be statistically distinguished (Figure S22G). To quantitatively understand these mRNA expression patterns, mathematical modeling was applied (Figure S23). Consistent with the previous analyses, a two-step model was required to fit *IL-1β* mRNA distributions in both the control (untreated with TSA) and TSA pre-treated cells (Figure 5F). In comparison to the lipid A control, the TSA pretreatment was associated with quantitative changes in kinetic parameter rates consistent with chromatin regulation (Figures 5G and S23B). In the first permissive step, the gene switching "on" rate was increased (from 0.007 to 0.012 min^{-1}, equivalent of a change in "off" time, from 142 to 83 min, for lipid A control versus lipid A +TSA, respectively) consistent with more frequent activation. Similarly, in the second step, TSA treatment also resulted in the increased "on" rate (from 0.05 to 0.1 min^{-1}, the equivalent of a change in "off" time from 20 to 10 min). No further changes were observed in other model parameters, although the transcription rate corresponding to the permissive step was reduced following TSA treatment (Figure S23B). Overall, these changes resulted in a significant quantitative increase in the moment estimates of the burst size and burst frequency following TSA treatment (Figure 5H). Overall, these suggest that regulation of the chromatin state may allow concurrent regulation of the burst size and frequency and thus modulation of the *IL-1β* gene-expression output .

DISCUSSION

While single-cell responses exhibit substantial cell-to-cell variability, a fundamental question remains how this variability is constrained. Here, we considered an endogenous TLR system with known differential responses to a range of immune-relevant conditions (Adamik et al., 2013; Escoubet-Lozach et al., 2011; Hao and Baltimore, 2009; Meissner et al., 2013; Ramirez-Carrozzi et al., 2009; Tong et al., 2016). Using our quantitative smFISH data as well as published scRNA-seq data (Shalek et al., 2014), we showed that variability in the expression of TLR-induced genes is constrained by gene-specific trends over a large range of mRNA expression (Figure 3). This demonstrates that the stimulation (or perturbation) merely modulates the variability of mRNA expression as a linear function of its mean. We further predicted that this theoretically imposes the constraints on the underlying transcriptional bursting characteristics in response to stimulation (Figure 4). We also demonstrated that in the case of bursty mRNA production, the expression variability (in terms of the mean-variance relationship) is essentially defined by the burst size, while responses to environmental changes are controlled via frequency modulation. In general, we predicted that both burst size and frequency may undergo modulation, with the contribution of the former increasing as the system departures from the bursty mRNA production regime. We validated these predictions using our *TNF-α* and *IL-1β* smFISH measurements as well as provided supporting evidence in a large set of TLR-regulated genes in primary immune cells (Figure 4). Finally, we demonstrated that modulation of chromatin state may account at least, in part, for the predicted modulation of transcriptional bursting of *IL-1β* expression (Figure 5).

We hypothesized that the observation that empirical mean-variance relationships are gene-specific, linear, and maintained over different conditions represent a fundamental property of a gene regulatory system. Our results imply an inverse relationship between the mean expression level and noise (Dar et al., 2016), while providing insight into the regulation of gene expression in response to stimulation. In agreement with previous work, we found that higher expression is accompanied by increased burstiness (Sanchez and Golding, 2013), while different levels of noise are associated with distinct subsets of parameter values (Figure S14). While not all the possible combinations of rates are ever explored in a biological system (Hausser et al., 2019), the fitted parameter changes (along with mean-variance trends) are dominated by modulation of gene activity switching rates, while modulation of transcription rates are associated with low transcriptional output (Figures S16 and S17). Our results perhaps reflect a related set of immune conditions used in the study (i.e., related ligands, dose-responses, co-stimulation, or generic chromatin perturbation), essentially affecting a large, well-connected signaling network (Oda and Kitano, 2006). As such, we found that these signals and conditions only modulate kinetic parameters of *TNF-α* and *IL-1β* transcription, rather than induce substantial changes in the mode of regulation. Recent analyses suggest that burst sizes are encoded within core promoters (Larsson et al., 2019). In fact, promoters of highly variable cytokine and chemokine genes are enriched for TATA boxes (Figure S3) and are depleted of CpG islands, in comparison to low heterogeneity TLR-dependent genes (Hagai et al., 2018). In turn, the frequency may be modulated via histone acetylation (Nicolas et al., 2018) or TF signaling events (Hagai et al., 2018; Wong et al., 2018). In the TLR system, the latter is likely related to the levels of upstream TFs or their activation patterns. For example, although heterogeneous, the NF-κB system activation exhibits dose-dependent and temporal regulation (Adamson et al., 2016; Bagnall et al., 2018; Muldoon et al., 2020; Selimkhanov et al., 2014; Sung et al., 2014; Wong et al., 2019), which might enable the fine-tuning of the underlying transcription and gene activation rates. It would be intriguing to understand whether gene-specific trends are sensitive to therapeutic compounds, or, in fact, pathogen stimulation. It would also be relevant to understand the modulation of parameter changes more broadly, i.e., in transcriptomics data, which in this work we only analyzed in terms of relative burstiness (Figure 4F). In particular, the linear mean-variance relationships theoretically imply a generic reciprocal relationship between the burst size and frequency (Figure 4), which would require further validation using more gene targets. While computationally feasible (Larsson et al., 2019), the present scRNA-seq dataset (Shalek et al., 2014) does not provide quantification of mRNA numbers, which is required to fit models accurately.

As a key part of this study, we quantitatively characterized the regulation of the TNF-α and IL-1 cytokines, which encode distinct roles during inflammatory responses and pathogen recognition (Lu et al., 2015). The expression of the short-lived

TNF-α mRNA transcript conforms to a simple one-step stochastic model of transcription (Raj et al., 2006; Skinner et al., 2016; Suter et al., 2011; Zenklusen et al., 2008), consistent with frequent transcription initiation but limited transcriptional output per burst (Figures 1 and 2). The transcription of *IL-1β* is characterized by lower bursting frequency and larger burst sizes compared with that of *TNF-α*. This behavior does not conform to a one-step model, and through mathematical modeling we showed that an intermediate regulatory step is required to explain the observed mRNA distributions. While previously considered complex models involved a promoter cycling step (Harper et al., 2011; Zoller et al., 2015), in our model, the intermediate step is associated with a low transcriptional output. This model structure is supported by the experimental evidence for a permissive step due to chromatin regulation. First, we observed biphasic patterns of *IL1α* and *IL1β* mRNA synthesis (Figure S21, which is also evident in previously published data; Shalek et al., 2014; Figure S5). Second, we found a marked temporal and spatial correlation between on-going *IL-1β* and *IL1α* transcription (as indicated by the co-localization of active transcription sites observed with smFISH; Figure S21), which likely underlies an association in the local chromatin structure (Iborra et al., 1998). Third, the chromatin modulator TSA altered the *IL-1β* mRNA distribution, resulting in more frequent mRNA bursts consistent with the two-step model (Figure 5). While elevating *IL-1β* expression, TSA treatment completely ablated the expression of *TNF-α* mRNA. This suggests that chromatin regulation may enable cytokine-specific control of the effector responses. In general, the co-association of multiple genes within common centers of mRNA synthesis provides an additional layer of regulation for gene expression, in which the combination of genes within an active factory might contribute synergistically to the timing, duration, and extent of synthesis from the spatially co-associated genes (Fanucchi et al., 2013; Li et al., 2012; Schoenfelder et al., 2010). The specific mechanisms involved in the regulation of the permissive chromatin states and robust *IL-1β* expression are not fully understood, but both cell-specific (e.g., heterogeneous signaling events) or cell-extrinsic (e.g., paracrine signaling) processes affecting TF-activation patterns (Lu et al., 2015; Shalek et al., 2013, 2014; Xue et al., 2015) might contribute to this. How many genes share complex modes of regulation, or, in general, whether functionally related genes exhibit co-variability in their responses is unclear. *IL1α* and *IL1β* represent one example of co-variability. It is currently unknown whether the heterogeneity of the *IL1β* gene and protein expression described here is fundamentally linked to the apparently cell-specific inflammasome activation as a mechanism to control cytokine levels in circulation and to minimize inflammasome-mediated cell death (Daniels et al., 2016). Further studies will also be required to quantitatively link the underlying TF dynamics, epigenetic control, and the target gene transcription, as well as protein expression and secretion (Junkin et al., 2016; Lee et al., 2014; Singer et al., 2014).

In summary, the study demonstrates that despite seemingly noisy responses, the heterogeneity of the single-cell and population-level TLR effector responses is defined by fundamental functional constraints. We propose that the constrained variability of the TLR-dependent gene response might be a key element of the antibacterial and inflammatory response and may constitute a common feature of inducible gene-expression systems in general.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Material Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Culture
  - Reagents
- METHOD DETAILS
  - Single-Cell RNA-seq
  - Single-Molecule RNA-FISH
  - Immunofluorescence
  - Experimental Design
  - Stochastic Modelling of Transcription
  - Model Fitting and Analysis
  - Noise Quantification in Count Data
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of scRNA-seq Data
  - Generic Properties of the TLR4 Response
  - smFISH Quantification
  - Point Estimators of Transcriptional Bursting
  - Modulation of Transcriptional Bursting
  - Burstiness in Genomics Data
  - Statistical Analyses

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cels.2020.08.007.

ACKNOWLEDGMENTS

We thank David Spiller and other members of the Systems Microscopy Centre in Manchester for discussions. We also thank Stuart Pepper for help with the Fluidigm system and the Bioimaging and Genomic Technologies Core Facilities in the Faculty of Biology, Medicine and Health in Manchester for assistance. We also thank Dale Norton for commenting on the manuscript as well as Conor Clerkin-Oliver, Ran Duan, Xinye Han, Zenthian Kai, Harriet Ninsiima, Stephen Richter, Dessislava Veltcheva, and Toby Wilkinson for the help with genomics analyses. This work was supported by BBSRC (BB/I017976/1, BB/K003097/1, and BB/R007691/1); Wellcome Trust (097820/Z/11/B and 220025/Z/19/Z); and by the National Science Center, Poland (https://ncn.gov.pl) grant 2016/23/B/ST6/03455. The work has received funding from the European Union Seventh Framework Programme (FP7/2012-2017) under grant agreement no. 305564.

AUTHOR CONTRIBUTIONS

Conceptualization, P.P., J.B., W.R., and M.M.; Methodology, J.B., W.R., N.A., C.C., M.M., and P.P.; Formal Analysis, J.B., W.R., N.A., J.R., and P.P.; Investigation, J.B., W.R., N.A., J.R., C.C., and P.P.; Resources, H.E.; Data Curation: J.B. and W.R.; Writing – Original Draft, P.P. Writing – Review & Editing, J.B., W.R., D.J., and M.M.; Visualization, J.B., W.R., N.A., and P.P.; Project administration, P.P.; Funding Acquisition, P.P., D.J., and M.M.; Supervision, M.P., M.M., and P.P.

DECLARATION OF INTERESTS

The authors declare no conflict of interests.

Received: April 12, 2019
Revised: April 8, 2020
Accepted: August 6, 2020
Published: September 11, 2020

REFERENCES

Adamik, J., Wang, K.Z.Q., Unlu, S., Su, A.J.A., Tannahill, G.M., Galson, D.L., O'Neill, L.A., and Auron, P.E. (2013). Distinct mechanisms for induction and tolerance regulate the immediate early genes encoding interleukin 1β and tumor necrosis factor α. PLoS One *8*, e70622.

Adamson, A., Boddington, C., Downton, P., Rowe, W., Bagnall, J., Lam, C., Maya-Mendoza, A., Schmidt, L., Harper, C.V., Spiller, D.G., et al. (2016). Signal transduction controls heterogeneous NF-κB dynamics and target gene expression through cytokine-specific refractory states. Nat. Commun. *7*, 12057.

Al-Mohy, A.H., and Higham, N.J. (2011). Computing the action of the matrix exponential, with an application to exponential integrators. SIAM J. Sci. Comput. *33*, 488–511.

Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H.B., Trombetta, J.J., Satija, R., Shalek, A.K., Xavier, R.J., Regev, A., and Hung, D.T. (2015). Pathogen cell-to-cell variability drives heterogeneity in host immune responses. Cell *162*, 1309–1321.

Bagnall, J., Boddington, C., Boyd, J., Brignall, R., Rowe, W., Jones, N.A., Schmidt, L., Spiller, D.G., White, M.R.H., and Paszek, P. (2015). Quantitative dynamic imaging of immune cell signalling using lentiviral gene transfer. Integr. Biol. (Camb.) *7*, 713–725.

Bagnall, J., Boddington, C., England, H., Brignall, R., Downton, P., Alsoufi, Z., Boyd, J., Rowe, W., Bennett, A., Walker, C., et al. (2018). Quantitative analysis of competitive cytokine signaling predicts tissue thresholds for the propagation of macrophage activation. Sci. Signal. *11*, eaaf3998.

Bagnall, J., Leedale, J., Taylor, S.E., Spiller, D.G., White, M.R., Sharkey, K.J., Bearon, R.N., and Sée, V. (2014). Tight control of hypoxia-inducible factor-alpha transient dynamics is essential for cell survival in hypoxia. J. Biol. Chem. *289*, 5549–5564.

Bradley, J.R. (2008). TNF-mediated inflammatory disease. J. Pathol. *214*, 149–160.

Bryant, C.E., Symmons, M., and Gay, N.J. (2015). Toll-like receptor signalling through macromolecular protein complexes. Mol. Immunol. *63*, 162–165.

Cheng, Z., Taylor, B., Ourthiague, D.R., and Hoffmann, A. (2015). Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation. Sci. Signal. *8*, ra69.

Daniels, M.J., Rivers-Auty, J., Schilling, T., Spencer, N.G., Watremez, W., Fasolino, V., Booth, S.J., White, C.S., Baldwin, A.G., Freeman, S., et al. (2016). Fenamate NSAIDs inhibit the NLRP3 inflammasome and protect against Alzheimer's disease in rodent models. Nat. Commun. *7*, 12504.

Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc. Natl. Acad. Sci. USA *109*, 17454–17459.

Dar, R.D., Shaffer, S.M., Singh, A., Razooky, B.S., Simpson, M.L., Raj, A., and Weinberger, L.S. (2016). Transcriptional bursting explains the noise-versus-mean relationship in mRNA and protein levels. PLoS One *11*, e0158298.

Dey, S.S., Foley, J.E., Limsirichai, P., Schaffer, D.V., and Arkin, A.P. (2015). Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol. *11*, 806.

Dinarello, C.A. (2011). Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. Blood *117*, 3720–3732.

Dueck, H., Khaladkar, M., Kim, T.K., Spaethling, J.M., Francis, C., Suresh, S., Fisher, S.A., Seale, P., Beck, S.G., Bartfai, T., et al. (2015). Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. Genome Biol *16*, 122.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183–1186.

Escoubet-Lozach, L., Benner, C., Kaikkonen, M.U., Lozach, J., Heinz, S., Spann, N.J., Crotti, A., Stender, J., Ghisletti, S., Reichart, D., et al. (2011). Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. PLoS Genet. *7*, e1002401.

Falvo, J.V., Tsytsykova, A.V., and Goldfeld, A.E. (2010). Transcriptional control of the TNF gene. Curr. Dir. Autoimmun. *11*, 27–60.

Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M.S., and Mhlanga, M.M. (2013). Chromosomal contact permits transcription between coregulated genes. Cell *155*, 606–620.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585–590.

Finch, W.H. (2012). Distribution of variables by method of outlier detection. Front. Psychol. *3*, 211.

Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. Science *315*, 972–976.

Gómez-Schiavon, M., Chen, L.F., West, A.E., and Buchler, N.E. (2017). BayFish: bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. Genome Biol. *18*, 164.

Hagai, T., Chen, X., Miragaia, R.J., Rostom, R., Gomes, T., Kunowska, N., Henriksson, J., Park, J.E., Proserpio, V., Donati, G., et al. (2018). Gene expression variability across cells and species shapes innate immunity. Nature *563*, 197–202.

Hao, S., and Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. Nat. Immunol. *10*, 281–288.

Harper, C.V., Finkenstädt, B., Woodcock, D.J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D.G., Mullins, J.J., Rand, D.A., Davis, J.R., and White, M.R.H. (2011). Dynamic analysis of stochastic transcription cycles. PLoS Biol. *9*, e1000607.

Hausser, J., Mayo, A., Keren, L., and Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. Nat. Commun. *10*, 68.

Hilfinger, A., and Paulsson, J. (2011). Separating intrinsic from extrinsic fluctuations in dynamic biological systems. Proc. Natl. Acad. Sci. USA *108*, 12167–12172.

Hu, X.Y., Chung, A.Y., Wu, I., Foldi, J., Chen, J., Ji, J.D., Tateya, T., Kang, Y.J., Han, J.H., Gessler, M., et al. (2008). Integrated regulation of toll-like receptor responses by Notch and interferon-gamma pathways. Immunity *29*, 691–703.

Iborra, F.J., Pombo, A., Jackson, D.A., and Cook, P.R. (1998). Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei (vol 109, pg 1427, 1996). J. Cell Sci. *111*, 2280.

Jackson, D.A. (2003). The anatomy of transcription sites. Curr. Opin. Cell Biol. *15*, 311–317.

Junkin, M., Kaestli, A.J., Cheng, Z., Jordi, C., Albayrak, C., Hoffmann, A., and Tay, S. (2016). High-content quantification of single-cell immune dynamics. Cell Rep. *15*, 411–422.

Larson, D.R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D.S., and Singer, R.H. (2013). Direct observation of frequency modulated transcription in single cells using light activation. eLife *2*, e00750.

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature *565*, 251–254.

Lee, R.E., Walker, S.R., Savery, K., Frank, D.A., and Gaudet, S. (2014). Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. Mol. Cell *53*, 867–879.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered

chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

Lu, Y., Xue, Q., Eisele, M.R., Sulistijo, E.S., Brower, K., Han, L., Amir, el-A.D., Pe'er, D., Miller-Jensen, K., and Fan, R. (2015). Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. Proc. Natl. Acad. Sci. USA *112*, E607–E615.

Martin, E.W., Pacholewska, A., Patel, H., Dashora, H., and Sung, M.H. (2020). Integrative analysis suggests cell type-specific decoding of NF-κB dynamics. Sci. Signal. *13*.

Martinon, F., Burns, K., and Tschopp, J. (2002). The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. Mol. Cell *10*, 417–426.

Medzhitov, R. (2007). TLR-mediated innate immune recognition. Semin. Immunol. *19*, 1–2.

Megaridis, M.R., Lu, Y.Y., Tevonian, E.N., Junger, K.M., Moy, J.M., Bohn-Wippert, K., and Dar, R.D. (2018). Fine-tuning of noise in gene expression with nucleosome remodeling. APL Bioeng. *2*, 026106.

Meissner, F., Scheltema, R.A., Mollenkopf, H.J., and Mann, M. (2013). Direct proteomic quantification of the secretome of activated immune cells. Science *340*, 475–478.

Molina, N., Suter, D.M., Cannavo, R., Zoller, B., Gotic, I., and Naef, F. (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. Proc. Natl. Acad. Sci. USA *110*, 20563–20568.

Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of transcripts in 3D FISH images. Nat. Methods *10*, 277–278.

Muldoon, J.J., Chuang, Y., Bagheri, N., and Leonard, J.N. (2020). Macrophages employ quorum licensing to regulate collective activation. Nat. Commun. *11*, 878.

Nicolas, D., Phillips, N.E., and Naef, F. (2017). What shapes eukaryotic transcriptional bursting? Mol. Biosyst. *13*, 1280–1290.

Nicolas, D., Zoller, B., Suter, D.M., and Naef, F. (2018). Modulation of transcriptional burst frequency by histone acetylation. Proc. Natl. Acad. Sci. USA *115*, 7153–7158.

Oda, K., and Kitano, H. (2006). A comprehensive map of the toll-like receptor signaling network. Mol. Syst. Biol. *2*, 2006.0015.

Paszek, P. (2007). Modeling stochasticity in gene regulation: characterization in the terms of the underlying distribution function. Bull. Math. Biol. *69*, 1567–1601.

Paszek, P., Ryan, S., Ashall, L., Sillitoe, K., Harper, C.V., Spiller, D.G., Rand, D.A., and White, M.R. (2010). Population robustness arising from cellular heterogeneity. Proc. Natl. Acad. Sci. USA *107*, 11644–11649.

Peccoud, J., and Ycart, B. (1995). Markovian modeling of gene-product synthesis. Theor. Popul. Biol. *48*, 222–234.

Raetz, C.R., Reynolds, C.M., Trent, M.S., and Bishop, R.E. (2007). Lipid A modification systems in gram-negative bacteria. Annu. Rev. Biochem. *76*, 295–329.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114–128.

Rhee, A., Cheong, R., and Levchenko, A. (2014). Noise decomposition of intracellular biochemical signaling networks using nonequivalent reporters. Proc. Natl. Acad. Sci. USA *111*, 17330–17335.

Sanchez, A., and Golding, I. (2013). Genetic determinants and cellular constraints in noisy gene expression. Science *342*, 1188–1193.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat. Genet. *42*, 53–61.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwarz, J., Schmidt, S., Will, O., Koudelka, T., Köhler, K., Boss, M., Rabe, B., Tholey, A., Scheller, J., Schmidt-Arras, D., et al. (2014). Polo-like kinase 2, a novel ADAM17 signaling component, regulates tumor necrosis factor α ectodomain shedding. J. Biol. Chem. *289*, 3080–3093.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. (2014). Systems biology. Accurate information transmission through dynamic biochemical signaling networks. Science *346*, 1370–1373.

Shahrezaei, V., and Swain, P.S. (2008). Analytical distributions for stochastic gene expression. Proc. Natl. Acad. Sci. USA *105*, 17256–17261.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Shen, H., Nelson, G., Nelson, D.E., Kennedy, S., Spiller, D.G., Griffiths, T., Paton, N., Oliver, S.G., White, M.R., and Kell, D.B. (2006). Automated tracking of gene expression in individual cells and cell compartments. J. R. Soc. Interface *3*, 787–794.

Sherman, M.S., Lorenz, K., Lanier, M.H., and Cohen, B.A. (2015). Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression. Cell Syst. *1*, 315–325.

Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol. Cell *55*, 319–331.

Skinner, S.O., Xu, H., Nagarkar-Jaiswal, S., Freire, P.R., Zwaka, T.P., and Golding, I. (2016). Single-cell analysis of transcription kinetics across the cell cycle. eLife *5*, e12175.

Smith, A.J., Keen, L.J., Billingham, M.J., Perry, M.J., Elson, C.J., Kirwan, J.R., Sims, J.E., Doherty, M., Spector, T.D., and Bidwell, J.L. (2004). Extended haplotypes and linkage disequilibrium in the IL1R1-IL1A-IL1B-IL1RN gene cluster: association with knee osteoarthritis. Genes Immun. *5*, 451–460.

So, L.H., Ghosh, A., Zong, C., Sepúlveda, L.A., Segev, R., and Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. Nat. Genet. *43*, 554–560.

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., 3rd, and Doyle, J. (2004). Robustness of cellular functions. Cell *118*, 675–685.

Sung, M.H., Li, N., Lao, Q., Gottschalk, R.A., Hager, G.L., and Fraser, I.D.C. (2014). Switching of the relative dominance between feedback mechanisms in lipopolysaccharide-induced NF-κB signaling. Sci. Signal. *7*, ra6.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472–474.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

Taylor, S.L., Renshaw, B.R., Garka, K.E., Smith, D.E., and Sims, J.E. (2002). Genomic organization of the interleukin-1 locus. Genomics *79*, 726–733.

Tong, A.J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. Cell *165*, 165–179.

Trapnell, C., Pachter, L., and Salzberg, S., L (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Vanhaecke, T., Papeleu, P., Elaut, G., and Rogiers, V. (2004). Trichostatin A-like hydroxamate histone deacetylase inhibitors as therapeutic agents: toxicological point of view. Curr. Med. Chem. *11*, 1629–1643.

Wong, V.C., Bass, V.L., Bullock, M.E., Chavali, A.K., Lee, R.E.C., Mothes, W., Gaudet, S., and Miller-Jensen, K. (2018). NF-κB-chromatin interactions drive diverse phenotypes by modulating transcriptional noise. Cell Rep. *22*, 585–599.

Wong, V.C., Mathew, S., Ramji, R., Gaudet, S., and Miller-Jensen, K. (2019). Fold-change detection of NF-kappa B at target genes with different transcript outputs. Biophys. J. *116*, 709–724.

Xue, Q., Lu, Y., Eisele, M.R., Sulistijo, E.S., Khan, N., Fan, R., and Miller-Jensen, K. (2015). Analysis of single-cell cytokine secretion reveals a role for paracrine signaling in coordinating macrophage responses to TLR4 stimulation. Sci. Signal. *8*, ra59.

Yagi, M., Miyamoto, T., Sawatani, Y., Iwamoto, K., Hosogane, N., Fujita, N., Morita, K., Ninomiya, K., Suzuki, T., Miyamoto, K., et al. (2005). DC-STAMP is essential for cell-cell fusion in osteoclasts and foreign body giant cells. J. Exp. Med. *202*, 345–351.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.

Zoller, B., Nicolas, D., Molina, N., and Naef, F. (2015). Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol. Syst. Biol. *11*, 823.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-IL1β	Abcam	Cat#Ab9722; RRID AB_308765
Anti-Rabbit (Alexa Fluor® 488 conjugated)	Abcam	Cat#Ab150077; RRID AB_2630356
Chemicals, Peptides, and Recombinant Proteins		
Lipid A Salmonella Minnesota Re595	VWR	Cat#CA80056-964
Recombinant mouse interferon-γ	Thermo Fisher	Cat#PMC4031
Dimethyloxalylglycine	Sigma-Aldrich	Cat#D3695
Trichostatin A	Sigma-Aldrich	Cat#T8552
Vectashield (with DAPI)	Vector Laboratories	Cat#H-1200
Critical Commercial Assays		
C1 Single-Cell Auto Prep IFC for mRNA Seq (10 – 17 mm)	Fluidigm	Cat#100-6041; RRID:IMSR JAX:000664
SMARTer Ultra Low RNA Kit for the Fluidigm C1 System	Clontech Takara	Cat#P/N 634833
Nextera XT DNA library preparation index kit	Illumina	Cat#FC-131-1002
Deposited Data		
scRNA-seq of RAW 264.7 cells to lipid A	This paper	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9219/
Codes for fitting mathematical models and scRNA-seq analyses, parameters of fitted models	This paper	https://github.com/ppaszek/transcriptionalBursting
scRNA-seq analysis of BMDC cells	(Shalek et al., 2014)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48968
Mouse reference genome NCBI build 38, GRCm38.p3	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/mouse/
Eukaryotic Promoter Database	EPD	https://epd.epfl.ch//index.php
Experimental Models: Cell Lines		
RAW264.7	ATCC	TIB-71; RRID CVCL_0493
Experimental Models: Organisms/Strains		
C57/BL (BMDM preparation)	Jackson Laboratory	JAX: 000664; RRID:IMSR JAX:000664
Oligonucleotides		
Probe sequence for smFISH (Table S2)	This paper	N/A
Software and Algorithms		
MATLAB	MathWorks	RRID:SCR_001622
FISH-QUANT	(Mueller et al., 2013)	https://www.bitbucket.org/muellerflorian/fish_quant/src/master/
GraphPad Prism 8	Graphpad software	RRID:SCR_002798
Cell Tracker	CellTracker software	https://www.warwick.ac.uk/fac/sci/dcs/people/till_bretschneider/celltracker/
Casava 1.8.3	Illumina	http://support.illumina.com/sequencing/sequencing_software/casava.html
Tophat 2.011	(Trapnell et al., 2009)	RRID:SCR_013035
Picard Tools	Piccard Tools software	RRID:SCR_006525
Matrix exponent algorithm	(Al-Mohy and Higham, 2011)	https://github.com/higham/expmv

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SoftWoRx 7.0	GE Healthcare	https://cdn.gelifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=17238
Stellaris RNA FISH Probe Designer	LGC Biosearch Technologies	https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pawel Paszek (pawel.paszek@manchester.ac.uk)

Material Availability

This study did not generate new materials

Data and Code Availability

The sequencing data generated during this study are available at ArrayExpress under accession no E-MTAB-9219 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9219/). MATLAB and Python codes generated during this study are available via Github [https://github.com/ppaszek/transcriptionalBursting]. The raw smFISH data including all the image files are too large to upload to existing public repositories, but these are available upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture

RAW 264.7 male murine macrophages (obtained from ATCC) were cultured in Dulbecco's modified eagle medium supplemented with 10% foetal bovine serum (Gibco) and 1% non-essential amino acids as described previously (Bagnall et al., 2015). Cells were not authenticated. Primary BMDMs were differentiated from bone marrow taken from the hind legs of adult 8-12 weeks male or female C57BL/6 mice (not involved in other procedures). Isolated cells were disrupted and homogenized by repeating pipetting until no lumps were visible. The cell suspension was then centrifuged at 200 g for 5 min and the resulting pellet re-suspended in DMEM (supplemented with 100 units/ml penicillin, 100 ug/ml streptomycin (all from Sigma-Aldrich, UK), 10% FCS (Gibco, UK), and 30% L929 cell-conditioned media) and then plated. After 72 h the media were replaced with fresh supplemented media. Cells were harvested (by washing with cold PBS) on day 6-8 and used for experiments within 24 h.

Reagents

Cells were stimulated with various doses of lipid A Salmonella Minnesota Re595 (VWR), 100ng/ml recombinant mouse interferon-γ (Life Technologies), 0.5mM dimethyloxalylglycine (Sigma-Aldrich) or 10μM Trichostatin A (Sigma-Aldrich). Slide mounting and nuclei staining was performed using Vectashield mounting medium with DAPI (Vector Laboratories).

METHOD DETAILS

Single-Cell RNA-seq

Single-cell sample collection and preparation was performed using the C1 Fluidigm platform, using the manufacturer's instruction. A suspension of appropriately stimulated $1x10^6$ RAW 264.7 cells per ml was prepared in serum-free media and appropriately mixed with C1 suspension reagent. The resulting cell mixture was then loaded into C1 Single Cell AutoPrep IFC microfluidic chip (calibrated for medium 10-17μm cell sizes). The microfluidic chip was then placed into the C1 Fluidigm system for processing, using the 'mRNA Seq: Cell Load' script. Verification of single-cell capture was performed by wide-field microscopy. Single-cell library construction was performed using the SMARTer Ultra Low RNA reagent kit (Takara®) for cDNA amplification, followed by the Nextera® XT DNA Index kit for fragmentation and barcoding of samples (Illumina®). DNA sequencing was performed by paired-end sequencing (100 + 100 cycles, plus indices) on an Illumina HiSeq2500 instrument.

Single-Molecule RNA-FISH

Custom Stellaris® FISH probes were designed against murine *IL1*α (NM_010554), *IL1*β (NM_008361) and *TNF*α- (NM_013693) cDNA sequences by utilising the Stellaris® FISH Probe Designer (Biosearch Technologies, Inc., Petaluma, CA). *IL1*α probes were conjugated with the Quasar-570 dye. *TNF*α and *IL1*β probes were conjugated with either the Quasar-570 dye or Quasar-670 dye for multiplexing with *IL1*α probes (see Table S2 for tabulated smFISH counts and probe lists). Cells were plated into 12-well plates

containing sterilised glass cover slips. After adherence, appropriately stimulated cells were fixed and labelled using Stellaris® protocol, following manufacturer's instructions (including co-immunofluorescence for protein levels). Samples were imaged using a DeltaVision (Applied Precision) wide-field microscope with a 60x/N.A.1.42 oil immersion Plan Apo N objective and Sedat Quad filter set was used. The images were collected using a Coolsnap HQ (Photometrics) camera with a z optical spacing of 0.2 μm.

Immunofluorescence

Cells were plated onto sterile glass coverslips submerged in media, and after adherence stimulated as required. Cells were fixed by immersion in 4% paraformaldehyde for 15 mins and then washed with PBS. Samples were incubated in the presence of 1:100 anti-IL1β primary antibody (abcam; ab9722) for 1 h at room temperature, washed and further incubated for 30 mins in the presence of 1:500 secondary antibody (abcam; ab150077) before a final PBS wash. The glass coverslip was then mounted on to a glass slide ready for imaging. Confocal microscopy was used to visualise anti-IL1β staining. FITC conjugates were excited using a 488 nm laser line and emitted signal detected after passing through a 505-550 nm bandpass filter, using LSM510 photomultiplier detectors. For quantitative comparison of fluorescence, all images were taken together using the same detector settings. Fluorescence levels were quantified using Cell Tracker Version 0.6 (Shen et al., 2006).

Experimental Design

smFISH data are representative of at least 2 biological replicates, scRNA-seq analysis of lipid A stimulated RAW 264.7 cells were performed using 1 replicate. Data were not randomized, stratified, or blinded for any of the analyses performed in this paper.

Stochastic Modelling of Transcription

CME Description

Temporal mRNA distributions for considered models of transcription are obtained using the Chemical Master Equation (CME) following approach by (Gómez-Schiavon et al., 2017). In brief, an infinite set of ordinary differential equations (ODEs) describes the flow of the probability in the biochemical system being in a particular state x and time t, $P(x,t)$ over all possible biochemical reactions k into and out of x:

$$\frac{dP(\boldsymbol{x},t)}{dt} = \sum_k [a_k(\boldsymbol{x}-v_k)P(\boldsymbol{x}-v_k,t) - a_k(\boldsymbol{x})P(\boldsymbol{x},t)]$$

$a_k \partial t$ denotes the probability that a biochemical reaction k will occur in the infinitesimal time interval ∂t, given that the system is in the state x, v_k is a stochiometric vector of reaction k that describes how the system changes when reaction k occurs. In general, CME is written in the matrix form as

$$\frac{dP(\boldsymbol{X},t)}{dt} = R(\theta)P(\boldsymbol{X},t),$$

where $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x_N}]^T$ is a vector of all possible cell states, $P(\boldsymbol{X},t) = [P(\boldsymbol{x}_1,t), P(\boldsymbol{x}_2,t), \ldots P(\boldsymbol{x_N},t)]^T$ and $R(\theta)$ is a transition rate matrix given by:

$$R_{ij}(\theta) = \begin{cases} -\sum_k a_k(\boldsymbol{x}_i) & \text{if } i=j \\ a_k(\boldsymbol{x}_j) & \forall j \text{ such that } \boldsymbol{x}_j = \boldsymbol{x}_i - \boldsymbol{v}_k \\ 0 & \text{otherwise.} \end{cases}$$

Time evolution of the probability distribution $P(\boldsymbol{X},t)$ is given by

$$P(\boldsymbol{X},t) = \exp[R(\theta)t]P_0(\boldsymbol{X}),$$

where $P_0(\boldsymbol{X})$ is specified by initial data that should satisfy $\sum_{\boldsymbol{X}} P_0(\boldsymbol{X}) = 1$. $P(\boldsymbol{X},t)$ is calculated using a fast matrix exponential function implemented in MATLAB by (Al-Mohy and Higham, 2011). All simulations begin with initial data in which no mRNA are present and both gene alleles in the '*off*' state. For practical purposes, the total number of mRNA molecules in the system—and hence the total number of states in the stochastic process—is truncated at $M = 2000$.

In general, $R(\theta)$ depends on both the model structure and the parameters. In this work, we considered a family of four transcriptional models of increasing complexity (as highlighted in Figure S11D). In the simplest model—often called the *telegraph model*—we assume two independent alleles for each gene, the activity of which switches randomly between '*off*' and '*on*' states, with only the latter being permissive for mRNA transcription (Raj et al., 2006; Skinner et al., 2016; Suter et al., 2011; Zenklusen et al., 2008). The associated kinetic parameters include switching '*on*' and '*off*' rates (k_{on} and k_{off}, respectively) as well as rates of mRNA transcription and degradation (k_t and k_d, respectively). The state of the cell in the telegraph model $x \in [s, m]^T$ is defined by the number of active alleles, s and number of mRNA molecules, m. The total number of states is $N = 3(M+1)$, subject to the constrains on the number of mRNA molecules M. A considered variant of the model includes an additional constitutive transcription rate k_0, which is incorporated into the transition matrix (see Figure S11D model 2).

We also consider an extension to the telegraph model that includes an additional regulatory step, which may be considered as a chromatin opening step that is required for full transcriptional activity. In the extended models, each allele exists in one of three states:

an inactive '*off*' state, an intermediate 'I' state or an active '*on*' state. Reversible stochastic transitions (with appropriate rates) occur between the inactive and intermediate as well as the intermediate and active states (but not directly between inactive and active states). We further assume that transcription occurs only in the active state (Figure S11D model 3) or in both the intermediate and active states (i.e. *IL1β* model, Figure S11D model 4). Given the upper bound on the number of mRNA molecules *M*, the total number of states in the extended models is $N = 6(M+1)$.

Model Fitting and Analysis

In order to investigate different regulatory scenarios (Figure S11B), we calculated exact temporal mRNA distributions using the CME approach as sketched above. A genetic algorithm (GA) was implemented using the *ga* function in MATLAB and employed to estimate model parameters, minimising the integrated absolute distance between the theoretical (CME) and measured cumulative distribution functions (CDFs). CDFs were fitted using *fitdist* function (with an Epanechnikov kernel function). The best 50 model fits from independent GA runs for each condition (using a population size of 200, elite count of 2, crossover factor of 0.6, and the tournament selection function). Gene activation rates were constrained to lie below 0.2 min^{-1}, while the degradation rate for *TNFα* transcripts was constrained to lie between 0.006 and 0.07 min (half-life between 10 and 115 mins), while the degradation rate for *IL1β* transcripts was constrained to lie between 0.002 and 0.006 min (half-life between 115 and ~350 mins). This is in an agreement with a short *TNFα* mRNA half-life (up to 1.5 h) in comparison to that of *IL1β* (stable at the time-scale of a 6 h experiment) measured in macrophages (Hao and Baltimore, 2009). We assumed two independent alleles per gene with the transcription rate constrained by 30 mRNA min^{-1} per allele. Rates as high as 2 to 10 mRNA min^{-1} have been reported for specific genes (Molina et al., 2013; Schwanhäusser et al., 2011; Skinner et al., 2016; Suter et al., 2011). In our dataset 10% of RAW 264.7 cells produced in excess of 200 mRNA h^{-1} (and 1% in excess of 400 mRNA h^{-1}), which is equivalent to a transcription rate between 1.67 to 3.33 mRNA min^{-1} per allele assuming constant production and no decay. Note that these are underestimates, as they assume steady production, while our transcription site data indicates intermittent transcriptional initiation.

The CME approach was also used to calculate sensitivity indexes corresponding to 10% parameter changes of the noise level $[1-\sigma_{10}/\mu_{10}/(\sigma_0/\mu_0)]$, where σ_0 and μ_0 correspond to nominal parameter values. Sensitivity indexes were calculated for distributions obtained at 180 mins after stimulation for one-step model for TNFα (Figure 2D), two-step *IL1β* model (Figure 2E) or one-step model refitted to recapitulate heterogeneity of *IL1β* expression (Figure S11F).

Noise Quantification in Count Data

Single-cell heterogeneity may emerge due to intrinsic stochastic fluctuations (i.e., random on-off switching) and extrinsic differences between cells (Elowitz et al., 2002; Hilfinger and Paulsson, 2011; Sherman et al., 2015). Therefore, in order to apply stochastic models of transcription (which assume intrinsic noise), we investigated the sources of the variability in the smFISH count data. Overall, these analyses suggest that intrinsic noise is a dominant factor in our datasets. In agreement, we show that one-step telegraph models explain all, but ~ 30% variability in data for *TNFα* smFISH counts (Figure 4B), while two-step model capture most of the variability in the *IL1β* data (Figure S18C).

(1) We demonstrate that count data exhibit intrinsic noise properties:
   (i) Noise decreases monotonically with mean μ in smFISH data (Figure S11A) as well as in our (Figure S3C) and published scRNA-seq (Figure S20E).
   (ii) In the limit of high μ, noise decreases sharply (Figure S11A), rather than approaching a plateau (Taniguchi et al., 2010).
   (iii) *TNFα* smFISH counts (Figure S16) as well as a majority of the scRNA-seq distributions (Figure S20A) fit negative binomial distributions. Of note, *IL1β* smFISH counts do not follow negative binomial distribution (Figure S16), since they represent a more complex model of regulation.

(2) A formal noise decomposition of the *TNFα* and *IL1β* dose-response count data (Rhee et al., 2014) shows that contribution of the intrinsic noise is dominant (across most condition), albeit also highlighted an extrinsic noise component (Figure S11B). To analyse potential sources of extrinsic noise, we show that
   (i) The percentage of variance of *TNFα* and *IL1β* smFISH counts explained by cell size (R^2 of the linear fit) is <7% (Figure S10B).
   (ii) The percentage of variance explained by regressing *TNFα* against *IL1β* counts is <20%, but ~41% for *IL1α* against *IL1β* (Figure S21). The former is consistent with extrinsic variability due to shared TLR signalling machinery, for example signalling dynamics (Wong et al., 2018, 2019), while the latter highlight the shared chromatin regulatory step.

(3) Currently, our smFISH datasets include between 10^2 and 10^3 individual cells per conditions (up to 18 conditions per probe) and up to 96 cells in scRNA-seq. In general, larger sample sizes might allow obtaining more accurate estimates of the underlying probability distribution function and their moments.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of scRNA-seq Data

Demultiplexing of the output data (allowing one mismatch) and BCL-to-Fastq conversion was performed with CASAVA 1.8.3. Reads were mapped to the *mus musculus* genome (assembly GRCm38.p3, downloaded from Ensembl) using Tophat version 2.011 (Trapnell et al., 2009) and assigned to genomic features in the corresponding gtf file using featureCounts in the Subread package (version

1-4.6). Counts for each gene were normalized to the median counts per cell (Figures S1A and S1B), data is presented as log_2*(normalised counts+1)* following analyses by (Shalek et al., 2013). 61 cells were included in the analysis with more than 2 million counts. A PCA plot of the normalized counts reveals a relatively uniform distribution of cells with no outliers or apparent overriding trends in the projection (Figure S1C). Comparison of normalized gene expression counts between two representative single cells shows a relatively strong correlation as observed in Figure S1D. Expression of housekeeping genes displays an almost linear correlation between these two cells (Spearman rank correlation of *0.94*). In contrast, genes whose expression was regulated in response to lipid A, display far greater variation (Spearman rank correlation of *0.71*). The mean of the transcript levels from the single cells were then compared to previously published population-level data performed under the same experimental conditions (Bagnall et al., 2015). These data were downloaded, re-mapped and re-normalised (as described here) to ensure parity between the datasets. A strong correlation between the population-level and single-cell data ($r=0.85$) was observed, confirming that library preparation preserved overall gene expression patterns (Figure S1E). As in Shalek et al (Shalek et al., 2013) we analysed the correlation between mean normalized counts (across all 61 cells) and the variability of these counts. We observe an inverse relationship between the normalized mean counts and the coefficient of variation (Figure S1F). All the housekeeping genes exhibit extremely low variability across the cells, while the lipid A responsive genes show far greater variability at comparable expression levels. Subsequently, a stringent cut-off was enforced to remove genes with high technical variability, leaving 1941 high-confidence genes. Data was clustered using the affinity propagation algorithm (Frey and Dueck, 2007); an unsupervised non-parametric method, which provides automated determination of numbers of clusters. Derived p-values were corrected for multiple testing using the method described by Benjamini and Hochberg (see Table S1 for normalised read counts and clustering analyses).

Generic Properties of the TLR4 Response

Eukaryotic promotor database (EPD) was used to determine TATA-box enrichment in the clusters displayed in Figure 1B. We observe significant enrichment of TATA sites in the promoter regions of the highly variable genes that failed to cluster (8 out of 10 genes have TATA boxes in their promoter regions, Figure S3A). In contrast, we do not find enrichment for TATA boxes in the promotor regions of the housekeeping genes examined. When comparing the variability in transcript levels of all genes within the (HC) single-cell dataset with and without TATA boxes, we find there is no statistical difference between the two sets (Figure S3B). In part this may be determined by the cut-off in transcript levels of high confidence genes, i.e. the HC is by definition less variable. Previously correlation has been found between variability in transcript level and mRNA half-life in a single cell study examining a variety of rat and mouse tissues (Dueck et al., 2015). A plot of the variability in expression levels of genes within this dataset and previously published mRNA half-lives (Schwanhäusser et al., 2011) reveals a limited negative correlation between half-life and heterogeneity, perhaps due to the fact that all recorded half-lives are large (e.g., >4 h, Figure S3B). Yet, there is a strong association between mRNA abundance (and transcript synthesis rate) and variation, indicative of a generic relationship between abundance and noise.

smFISH Quantification

Raw images were deconvolved using the SoftWoRx 7.0 software (GE Healthcare). Spot counting for mature and nascent mRNA was performed with FISH-Quant v2d (Mueller et al., 2013). The total cell area was calculated by extracting the number of pixels and pixel size in each drawn cell boundary. The nuclear area was calculated by applying the MATLAB function 'greythresh' to the maximum projection of the deconvolved DAPI signal. Pixel areas for each nuclear mask were extracted and scaled to the actual pixel sizes. For cell size normalisation, each individual mRNA count was scaled via the ratio of the average nuclear area of the population and nuclear area of the cell (Figure S10).

Point Estimators of Transcriptional Bursting

Transcriptional burst size and burst frequency are defined as the average number of mRNA produced per gene activation event, and the frequency of gene activation events, respectively (Nicolas et al., 2017). In the case of the one-step telegraph model, these are directly related to the kinetic parameters of transcription (Nicolas et al., 2018). When accounting for two independent alleles of a gene, in the steady-state burst size is defined as $b_k=k_t/k_{off}$, while bursting frequency is given by $f_k=2k_{on}k_{off}/(k_{on}+k_{off})/k_d$ (we refer to these as *kinetic estimators*). Alternatively, estimators based on the sample variance σ^2 and the mean μ of the mRNA distribution (referred to here as *moment estimators*) such that the burst size ($b_m=\sigma^2/\mu$) and burst frequency [$f_m=\mu/(b_m-1)$] are sometimes used (Nicolas et al., 2017; Raj et al., 2006; Suter et al., 2011; Wong et al., 2018). In general, moment estimators are used to describe burstiness, i.e. quantitative departure from a *non-bursty* (Poissonian) mRNA production (where $b_m=1$ *and* $f_m=\infty$) (Nicolas et al., 2017; So et al., 2011; Wong et al., 2018). To evaluate the difference between estimators we define an error function:

$$Error = \frac{kinetic\ parameters\ estimate - moment\ estimate}{kinetic\ parameters\ estimate}$$

Given expressions for the steady-state mRNA moments in the telegraph model (Peccoud and Ycart, 1995; Paszek, 2007; Shahrezaei and Swain, 2008), when accounting for two alleles we have that

$$\mu = \frac{2k_{on}k_t}{k_d(k_{off}+k_{on})} \qquad \text{(Equation 1)}$$

$$\sigma^2 = \frac{\mu^2}{2}\frac{k_{off}}{k_{on}\left(1+\frac{(k_{off}+k_{on})}{k_d}\right)} \quad \text{(Equation 2)}$$

In this work, we use moment estimators calculated either for smFISH and scRNA-seq data or for theoretical mRNA distributions (at any arbitrary time) obtained from the CME. For application of the one-step telegraph model we utilise kinetic parameter estimators.

Therefore, errors in the steady-state may be expressed as

$$b_{error} = 1 - \frac{{k_{off}}^2}{k_d(k_{off}+k_{on})+(k_{off}+k_{on})^2} - \frac{k_{off}}{k_t}, \quad \text{(Equation 3)}$$

$$f_{error} = 1 - \frac{(k_{off}+k_{on}+k_d)(k_{off}+k_{on})}{{k_{off}}^2}. \quad \text{(Equation 4)}$$

In general, the error associated with moment estimators depends on specific parameter values and the error in the bursting frequency is independent from the transcription rate k_t. As already well established in the literature (Nicolas et al., 2017), in the 'bursting' regime, corresponding to short and infrequent activation events, i.e. $k_{off}>>k_{on}$ and $k_d>>k_{off}$, errors resulting from moment estimators are negligible (given that in general $k_t>>k_{off}$), Equations 3 and 4. In this case both errors converge to 0, and thus moment estimators are as accurate as kinetic parameter estimators. In order to understand the generic suitability of moment estimators, we calculated errors associated across a wide range of parameter values using the fitted TNFα model (Figure S6). In the physiological parameter range, i.e. $k_{on}<0.1\ min^{-1}$ and $k_{off}<0.2\ min^{-1}$, and assuming $k_t<30\ mRNA/min^{-1}$, both errors are constrained ($f_{error}<1$ and $b_{error}<1$) for $k_{off}>k_{on}$ (Figure S6A). In the case of TNFα, where k_{off}/k_{on} is equal to 6, the errors due to approximation via moment estimators are ~30% (see also Figure S11G). These errors substantially increase when $k_{off}/k_{on}\sim1$, but are independent of transcription rate (at least above 5 mRNA/min, Figure S6B). Both errors also depend on the mRNA half-life, but within the physiological range, i.e. $k_d<0.01\ min^{-1}$ the corresponding changes are limited (for $k_{off}/k_{on}>3$). While kinetic estimators define bursting characteristics only at the steady-state, the moment estimators can be calculated at any time (Figure S6C). Temporal relationships (calculated based on theoretical distribution at 1, 3 and 6h) converge to the steady-state approximation (Figure S6A).

Modulation of Transcriptional Bursting

We theoretically calculated relationships between parameters of the telegraph model that satisfy empirically observed linear mean-variance relationships. We assume that the sample mean and variance of the gene expression distribution follows a general linear trend,

$$\sigma^2 = \alpha\mu + \sigma_0. \quad \text{(Equation 5)}$$

Under steady-state assumption, i.e. by using Equations 1 and 2, with $\sigma_0=0$, this relationship corresponds to

$$k_t = (\alpha-1)(k_{off}+k_{on}+k_d)\left(1+\frac{k_{on}}{k_{off}}\right), \quad \text{(Equation 6)}$$

whereas in general ($\alpha_0\neq0$):

$$\frac{k_{off}k_{on}k_t^2}{k_d(k_d+k_{off}+k_{on})(k_{off}+k_{on})^2} + (1-\alpha)\frac{k_{on}k_t}{k_d(k_{off}+k_{on})} - \frac{\sigma_0}{2} = 0. \quad \text{(Equation 7)}$$

The above equations define the relationship between kinetic parameters that satisfy linear constraints. Equation 6 describes a surface in the three-dimensional (k_{off}, k_{on}, k_t) parameter space on which the $\sigma^2=\alpha\mu$ relationship holds. We plot this surface, as well as bursting characteristics on the surface for $\alpha=100$ and $\alpha=10$ (corresponding to genes with different level of variability) for biologically plausible set of parameters, i.e. $k_{off}<0.2\ min^{-1}$ and $k_{on}<0.1\ min^{-1}$, while assuming $k_d=0.014\ min^{-1}$ (i.e., fitted *TNF*α degradation rate) and $k_t<30\ min^{-1}$ (Figures S14A–S14C).

To maintain a linear mean-variance relationship the system can move freely between different (k_{off}, k_{on}, k_t) parameter values, which results in modulation of both burst size and frequency. In the case of bursty expression regime, i.e. $k_{off}>>k_{on}$ (for $k_{off}>>k_d$) it follows from Equation 6 that $b_k=\alpha-1$ and $f_k=\mu/(\alpha-1)$. Therefore, burst size is necessarily constant (and equal to the slope of the mean-variance line for large α) over the range of mean mRNA response, while changes of gene expression are controlled solely by frequency modulation. If the system is not in the bursty regime, the extent of burst size and frequency modulation is related to the activation rate k_{off} (or in general k_{off}/k_{on} ratio, Figure S15). Based on the (k_{off}, k_{on}, k_t) parameter surfaces, the relationship between bursting frequency and relative change of the burst size was calculated for a range of regression slopes from α=10 to 200, (Figures S15A and S15B). We find that the larger k_{off}, the smaller are changes of the burst size, and in turn the larger are changes of the bursting frequency over the corresponding change of mean mRNA expression (Figure 4A).

When the system approaches a bursty regime, i.e. $k_{off}>>k_{on}$ the changes of burst size become negligible (Figure S15C). In this case, moment estimators provide an accurate description of the bursting characteristics for the one-step telegraph model (i.e., for $\alpha>>1$ moment and kinetic estimators are the same).

The generic case of non-zero intercept, i.e. $\sigma^2= \alpha\mu+\sigma_0$ is considered in Equation 7, where parameter regions consist of two roots of the quadratic equation (see Figures S14D and *S14*E). One of these roots overlaps with the solution of Equation 6 (the case of non-zero intercept), while the second, associated with a small transcription rate (Figure S14E) disappears as $\sigma_0 \rightarrow 0$. In the bursty regime, Equation 7 can be re-arranged as

$$f_k b_k^2 + (1-\alpha) f b - \sigma_0 = 0.$$

Given that $\mu=fb$, we have that

$$b_k = (\alpha - 1) + \frac{\sigma_0}{\mu}$$
$$f_k = \frac{\mu}{b_k}. \qquad \text{(Equation 8)}$$

Equation 8 show that for $\sigma_0 \neq 0$ the burst size is a non-monotonic function of the mean μ that diverges as μ tends to zero. When $\mu>>0$ the burst size tends asymptotically to the constant value $b_k = (\alpha - 1)$, so that the description is equivalent to using moment estimators, in the sense that the burst size is predetermined by the slope of the mean-variance line. Of note, for $\sigma_0>0$ burst size relationship has a minimum for $\mu=\sigma_0/(\alpha-1)$] and burst size monotonically increase (and *vice versa* for $\sigma_0<0$, Figures 4C and 4D). In this case, the description is equivalent to using moment estimators, such that the burst size is predetermined by the slope of the mean-variance line, and constant in the case of the zero intercept, i.e. $b_m=\sigma^2/\mu=\alpha+\sigma_0/\mu$, *while* for $\alpha>>1$, $b_m= b_k$. In addition, the frequency undergoes modulation as a function of the mean, i.e. $f_m=\mu/(b_m-1)=\mu/[(\alpha-1)+\sigma_0/\mu=f_k]$.

The comparison between moment (and kinetic estimators) is depicted in Figure 4C (in the case of a non-zero intercept for the fitted *TNF-α* smFISH data). The burst and frequency relationships are predicted based on coefficients of the linear mean-variance relationships (Figure 4B). In the case of *IL1β* (where the complexity of the model prevents analytical solutions), we use moment estimators based on the fitted smFISH dataset (Figure S18C). We find that while frequency changes are predicted accurately, the burst size is predicted accurately only for large means (Figure 4D). We find that specifically in the case of positive intercept (e.g., in the case of *IL1β*) the simple relationship does not reproduce the non-constant behavior at low mRNA levels. For the analyses of scRNA-seq datasets (Figures 4F and S19) we therefore fitted individual relationships separately (i.e. mean-variance, mean- burst-size, mean-frequency, etc.), rather than compare data with relationships predicted by the mean-variance line [Equation 8]. However, we then demonstrate that characteristics predicted by the theory are present in the fitted data, specifically there is a reciprocal relationship between burst size and frequency across considered genes (Figure S19F).

Intuitively, mean-variance relationships are expected to have zero intercepts. However, in both smFISH and scRNA-seq datasets we find evidence for both negative and positive intercepts. In the case of *TNFα* (negative intercept, Figure 4C), theoretical predictions of burst size and frequency based on the regression fit are consistent with fitted data and indeed predict a minimum in frequency changes. We find that in RAW 264.7 cells, there is always a basal (and substantial) expression of *TNFα* mRNA in unstimulated cells, which perhaps contributes to this behaviour (i.e. no true zero in the system). In general, fitted intercepts have relatively small values (comparing to the overall variance) and tend to be positive. This suggests elevated level of variance consistent with measurement noise (especially for small means). We accept that only a limited amount of data is available to be fitted per condition, thus individual fits may be affected by specific values of individual or groups of points. We consider these mean-variance relationships are empirical, and treat them as such in the manuscript.

Burstiness in Genomics Data

Inference of mean-variance relationship was performed using a dataset from BMDCs incorporating 29 scRNA-seq experiments (each corresponding to a single Fluidigm C1 experiment with up to 96 cells) on the response time-course (at 0, 1, 2, 3 and 6 h) as well as additional perturbations such as treatment with IFNβ, inhibition of paracrine secretion (chemical or physical on chip) or cell knockout for IFNR1 in and STAT1 expression (Shalek et al., 2014). We considered 812 genes that were induced by at least two-fold (compared to unstimulated cells) at the population level at any time point during the LPS stimulation [as identified in (Shalek et al., 2014)]. Visual inspection of the data revealed outliers in the linear regression fit, therefore, outlier removal method with Mahalanobis distance was used (with *0.05* threshold for outlier detection) (Finch, 2012). After removing low abundant genes (maximum mean expression <100 read counts) this resulted in 290 genes for the core TLR dataset (time-course) and 323 for the combined dataset (including perturbations). Bursting characteristics (based on moment estimators) for individual data points were fitted using linear regression and power functions (in semi log scale) when appropriate and presented as smooth curves. Robust regression (excluding data points with corresponding fitted residuals > $1.5 \cdot \sigma_{residuals}$) was used to either remove noisy data (as expected in the scRNA-seq measurement) or individual datapoints that did not affect the overall trend. Equations, fitted parameters, corresponding correlation coefficients and highlighted outliers are included in the Tables S3, S4, and S5. Fitting protocols were implemented in Python using R kernel, individual gene graphs were produced in MATLAB R2014a.

In order to validate estimates from scRNA-seq data, raw BAM dataset from Shalek et al. corresponding to LPS stimulation at 4h was downloaded and re-mapped using Picard Tools to remove duplicate reads (http://broadinstitute.github.io/picard/). Mapped

data was normalised to read counts per million and compared with the original dataset (Figure S20). Specifically, for the set of LPS-dependent genes characterised by linear mean-variance relationships, mean and variance, as well as relative burst size and frequency (based on moment estimators) were calculated. In addition, chi-squared goodness-of-fit tests was performed to determine whether count data (in each dataset) follow negative binomial distribution. p-values were adjusted using Benjamini-Hochberg procedure for false discovery rate, genes with <10 non-zero reads (out of 95 captured cells) were not considered.

Statistical Analyses

Data are described by the sample mean and standard deviation (SD). Sample size are provided ion figure legends. All statistical analyses were performed in GraphPad Prism 8 or MATLAB. Data were checked for normality with the D'Agostino-Pearson omnibus test. When normal, parametric tests were performed (t-test, standard one-way ANOVA); otherwise, non-parametric tests are used (Mann-Whitney, Kruskal-Wallis ANOVA). Tukey's or Dunn's correction for multiple comparisons was applied, respectively. Contingency tables were assed with Fisher exact tests. MATLAB's *chi2gof* chi-squared goodness-of-fit test was performed between count distributions and respective negative binomial distributions with parameters estimated from the data (using *fitdist* function). Benjamini and Hochberg method was used for multiple comparison adjustment in genomics data. Significance was defined for p-value (and adjusted p value, when relevant) <0.05. Details of all statistical tests are provided in the corresponding figure legends.

Cell Systems, Volume *11*

Supplemental Information

Gene-Specific Linear Trends Constrain Transcriptional Variability of the Toll-like Receptor Signaling

James Bagnall, William Rowe, Nissrin Alachkar, James Roberts, Hazel England, Christopher Clark, Mark Platt, Dean A. Jackson, Mark Muldoon, and Pawel Paszek

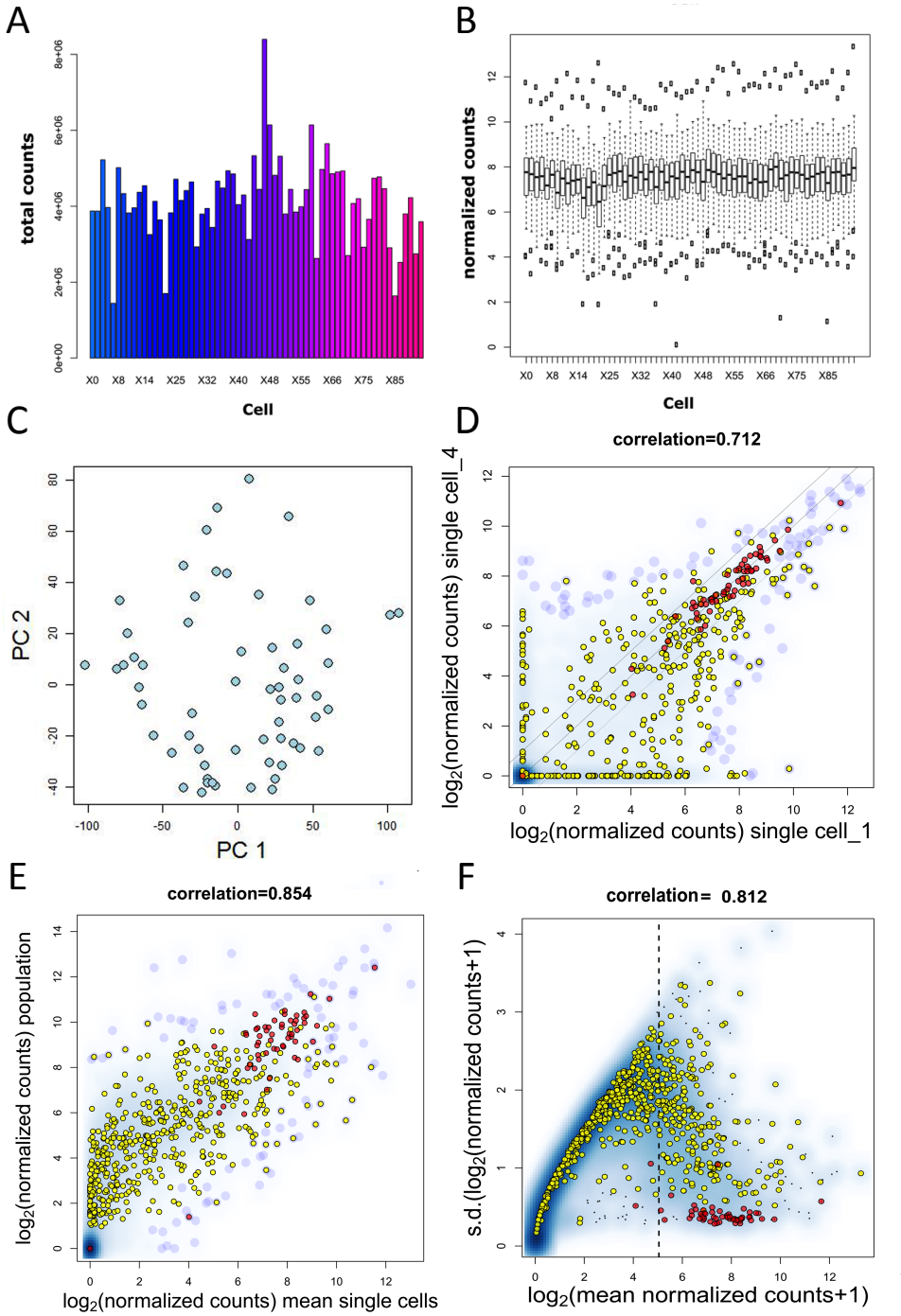


Figure S1. Variability within the scRNA-seq data, related to Figure 1. A. Total read counts across 64 cells before normalisation. 61 cells with read counts *>2x10*^{6} selected for subsequent analysis. **B**. Box-plots displaying the distribution of expression levels (in $\log_2$ scale) of housekeeping genes across single cells post normalisation by median count per cell (*x10*). **C**. PCA plot displaying variability between single-cell expression levels post normalisation. **D.** Normalized gene expression counts between two representative single cells. Shown is the smooth scatter plot, including lipid A response (in yellow), housekeeping genes (in red) and

other high confidence (HC) genes (in blue). Spearman rank coefficient indicates correlation between all genes between two cells. **E**. Comparison between transcript levels from population-level (from [1]) and mean of single cells. Shown is the smooth scatter plot, including lipid A response (in yellow), housekeeping genes (in red) and other genes (in blue). Spearman rank coefficient indicates correlation between all genes in the population (from [1]) and mean of single cells. **F.** Smooth scatter plot of standard deviation (SD) vs. mean normalized data. High confidence (HC) gene set defined for the expression level above cut-off line of $log_2(counts+1)=5$. Spearman rank coefficient indicates correlation between SD and mean normalized data (for all genes). Genes colour-coded as in D.

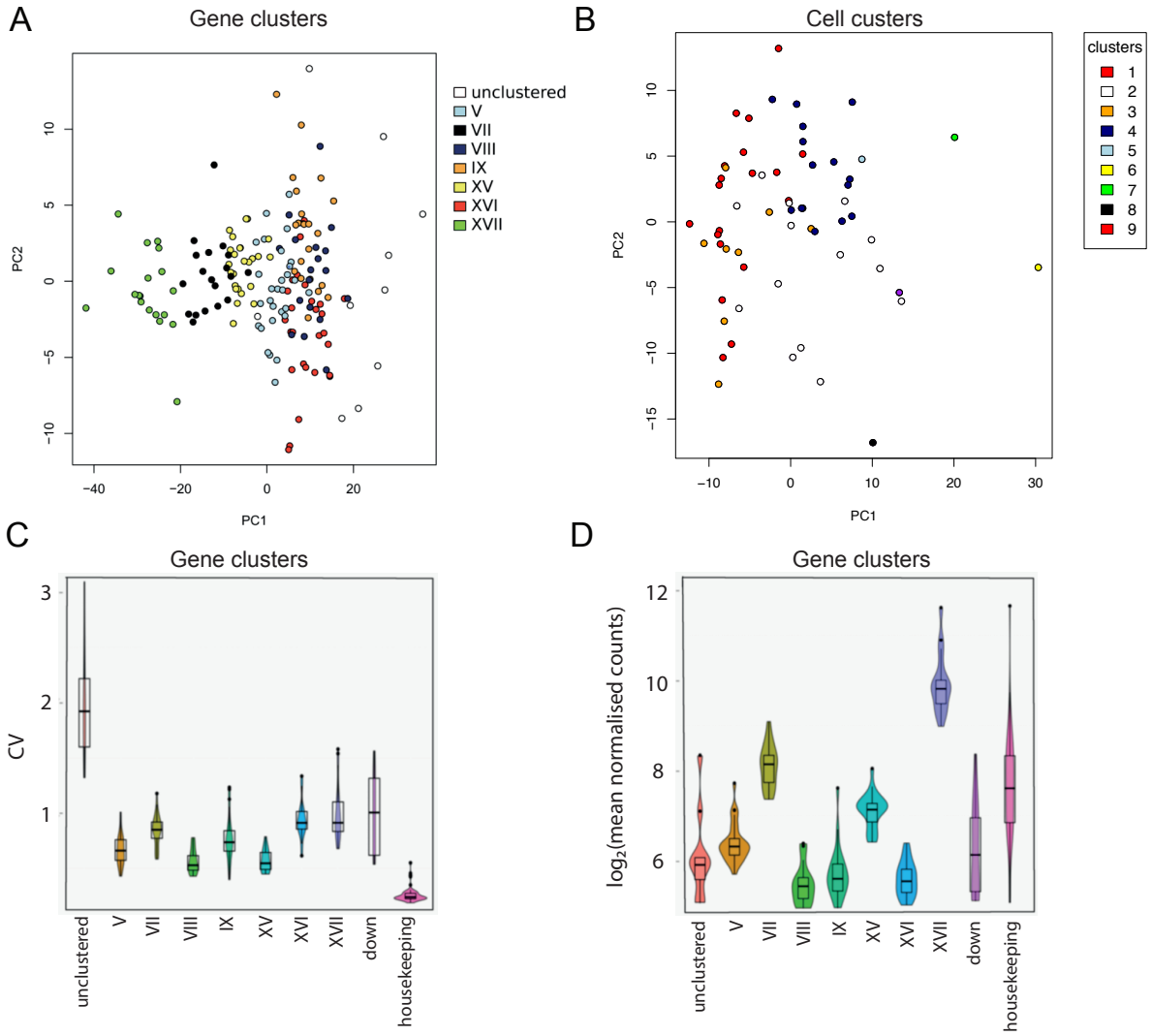


Figure S2. Clustering analysis of scRNA-seq data on lipid A-stimulated RAW 264.7 macrophages, related to Figure 1. A. Principal component analysis (PCA) of gene clusters from Fig. 1B. Shown are the first two principal components (PC1 vs. PC2). **B**. Principal component analysis of cell clusters identified in Fig. 1B. **C.** Violin plots displaying coefficient of variation (CV) of normalised transcript levels across gene clusters. **D.** Violin plots displaying mean normalized transcript levels across gene clusters.

A

group	TATA-box	total	p.value
un-clustered	8	10	<0.0001
V	3	28	1
VII	7	19	0.0045
VIII	2	17	1
IX	3	9	0.0793
XV	7	35	0.1784
XVI	6	27	0.1228
XVII	6	18	0.0143
Down-regulated	2	12	0.6424
housekeeping	15	55	0.0023
High-confidence set	227	1942	

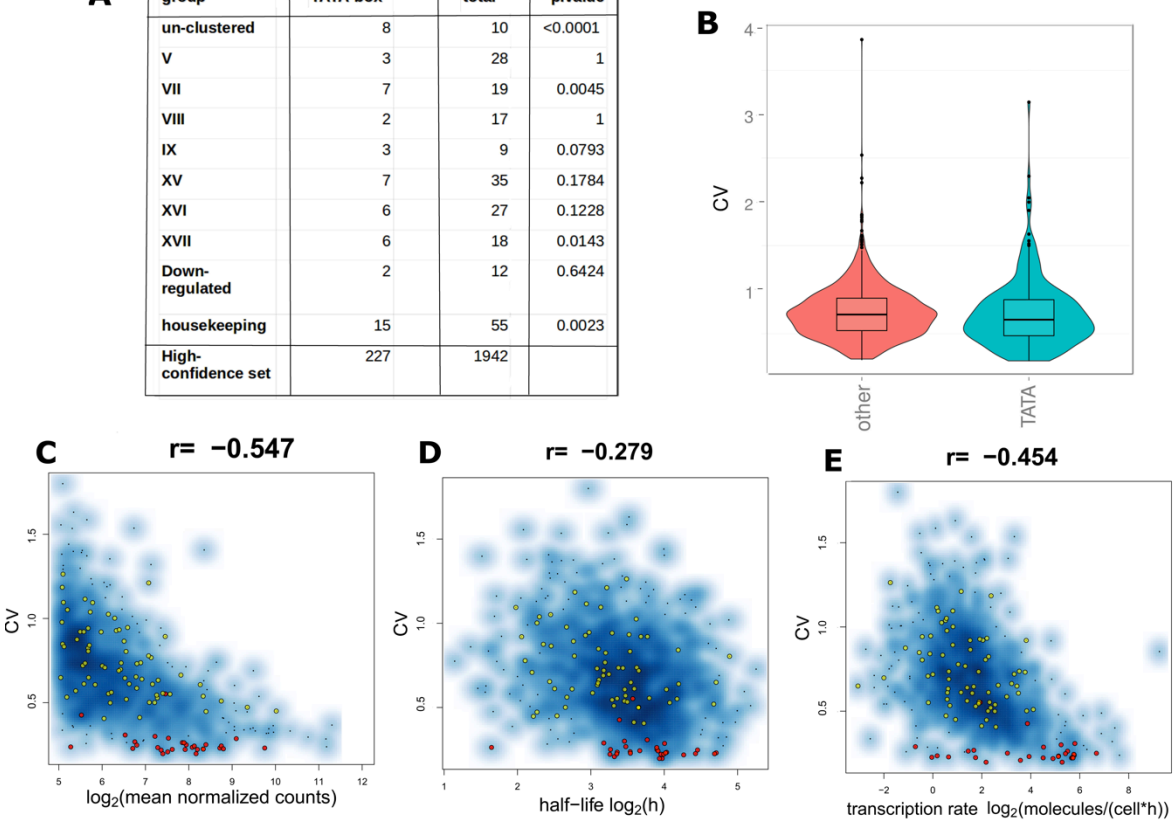


Figure S3. Physical properties of genes contribute to variation in gene expression measured in scRNA-seq data, related to Figure 1. A. Enrichment of TATA boxes the upstream regions of genes from clusters in Fig. 1B. **B.** Boxplot displaying coefficient of variation of genes with and without (proximal to) TATA-boxes (within the HC group). **C-E**. Coefficient of variation of individual genes against mean normalised transcript levels (C), mRNA half-life (D) and rate of transcription (E) (values taken from [2]). Shown are smooth scatter plots, including lipid A response (in yellow), housekeeping genes (in red) and other high confidence (HC) genes (in blue). Correlation coefficients calculated for all genes, assessed as statistically significant (p-val <0.001) according to Spearman rank correlation test.

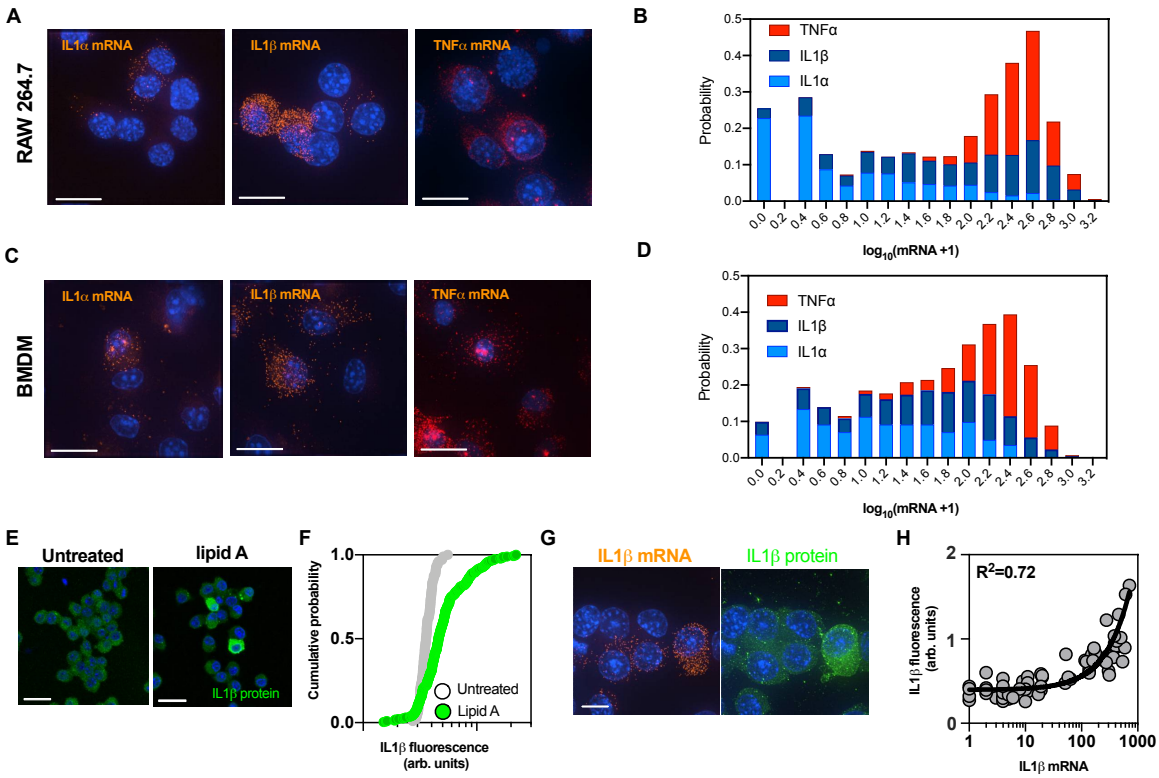


Figure S4. Analysis of TNFα, IL1α and IL1β expression, related to Figure 1. A. smFISH analysis of *IL1α*, *IL1β* and *TNFα* mRNA expression in RAW 264.7 cells. Shown are maximum intensity projections from deconvolved wide-field microscopy image z-stacks of representative cells stimulated with 500ng/ml lipid A for 3 h. mRNA transcript shown in orange, DAPI nuclear staining depicted in blue. Scale bar 10 μm. **B.** Distribution of mRNA counts from A. Shown are histograms for the *IL1α*, *IL1β* and *TNFα* abundances expressed as $\log_{10}$(mRNA+1) across at least three replicates, from 447, 718 and 356 cells, respectively. **C.** smFISH analysis of *IL1α*, *IL1β* and *TNFα* mRNA expression in BMDM cells. Shown are maximum intensity projections from deconvolved wide-field microscopy image z-stacks of representative cells stimulated with 500ng/ml lipid A for 3 h. mRNA transcript shown in orange, DAPI nuclear staining depicted in blue. Scale bar 10 μm. **D.** Histogram of mRNA counts in BMDMs (from C). Shown is analysis of 142, 732 and 322 cells for *IL1α*, *IL1β* ad *TNFα* across three replicates, respectively. **E**. Immunostaining of *IL1β* protein expression in RAW 264.7 macrophages. Shown are confocal microscopy images of cells treated with 500 ng/ml of lipid A for 3 h (or untreated). *IL1β* protein shown in green, DAPI nuclear staining depicted in blue. Scale bar 20 μm. **F**. Cumulative probability distribution of *IL1β* protein from A. Shown is analysis of 104 and 316 of untreated and lipid A-treated cells across three replicates, respectively. **G**. Dual smFISH and immunostaining analysis of *IL1β* mRNA and protein levels. Shown are deconvolved wide-field

microscopy images of representative RAW 264.7 cells stimulated with 500 ng/ml of lipid A for 3 h. mRNA transcript shown in orange (left panel), protein in green (right panel), DAPI nuclear staining depicted in blue. Scale bar 10 μm. **H**. Correlation between *IL1β* mRNA and protein levels in cells from C. Shown are individual cell counts depicted with circles, in black a nonlinear regression fit (with a coefficient of determination R^2).

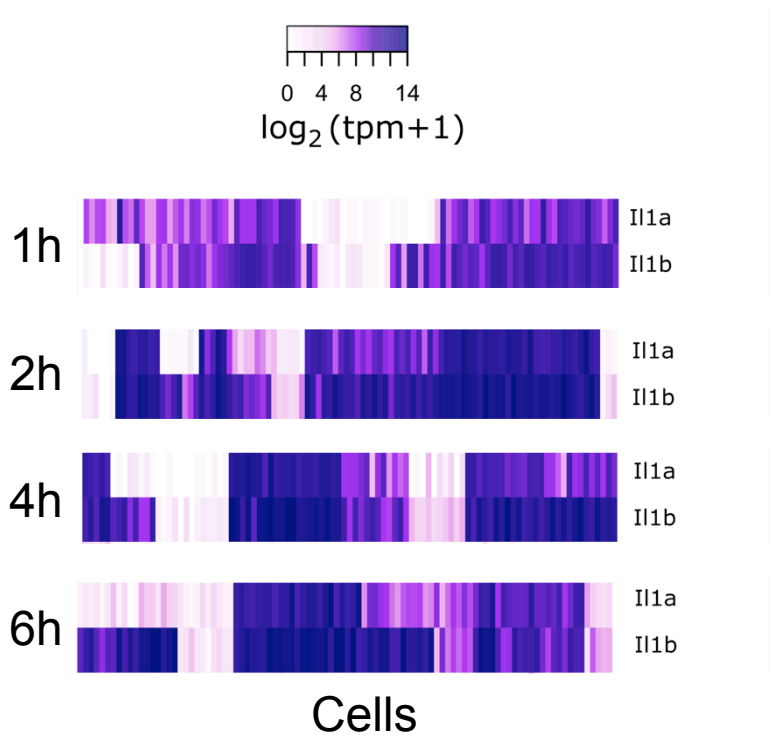


Figure S5. Analysis of *IL1α* and *IL1β* expression in mouse bone marrow derived dendritic cells, related to Figure 1. Heat maps displaying single cell expression of *IL1α* and *IL1β* across cells. Transcript levels measured as transcripts per million (tpm) were downloaded from the supplementary data from (Shalek et al., 2014). Expression levels are shown at 1, 2, 4 and 6 hours after LPS stimulation.

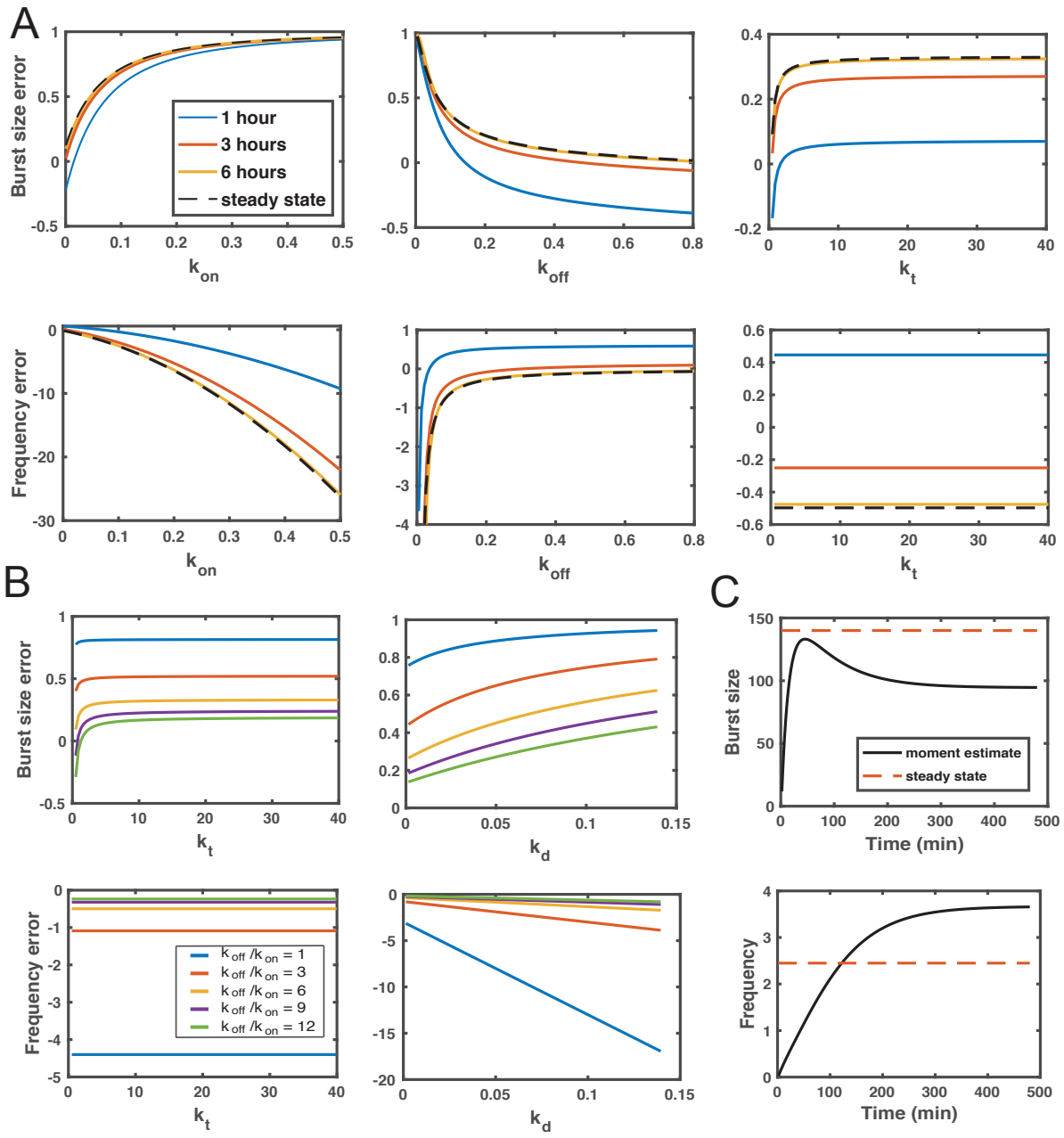


Figure S6. Comparison of point estimators of transcriptional bursting parameters, related to Figure 2. A. Fractional errors of moment estimators as a function of kinetic parameters. Shown are simulations using parameters of the fitted *TNFα* telegraph model (k_{on}=0.02 min^{-1}, k_{off}=0.12 min^{-1}, k_t=16.8 mRNA/min, k_d=0.014 min^{-1}) for systematic changes of individual parameter values (k_{on}, k_{off} and k_t, while retaining values of other). In colour lines, theoretical moment estimators using exact temporal mRNA distributions at 1, 3, 6 h; in broken lines steady-state errors given by Eqs (3) and (4). **B.** Steady-state burst size and frequency errors as a function of the 'burstiness'. Errors calculated for the parameters of the fitted TNFα model (k_{on}=0.02 min^{-1}, k_{off}=0.12 min^{-1}, k_t=16.8 mRNA/min, k_d=0.014 min^{-1}) for systematic changes of transcription (k_t) and degradation (k_d) rates. 'burstiness' defined as k_{off}/k_{on} ratio and

simulated by changing k_{off} rate (as highlighted), while maintaining k_{on} constant. **C.** Estimates of burst size and frequency for the fitted *TNFα* model at different time points (k_{on}=0.02 min^{-1}, k_{off}=0.12 min^{-1}, k_t=16.8 mRNA/min, k_d=0.014 min^{-1}). In broken lines are steady-state estimates using fitted kinetic parameters.

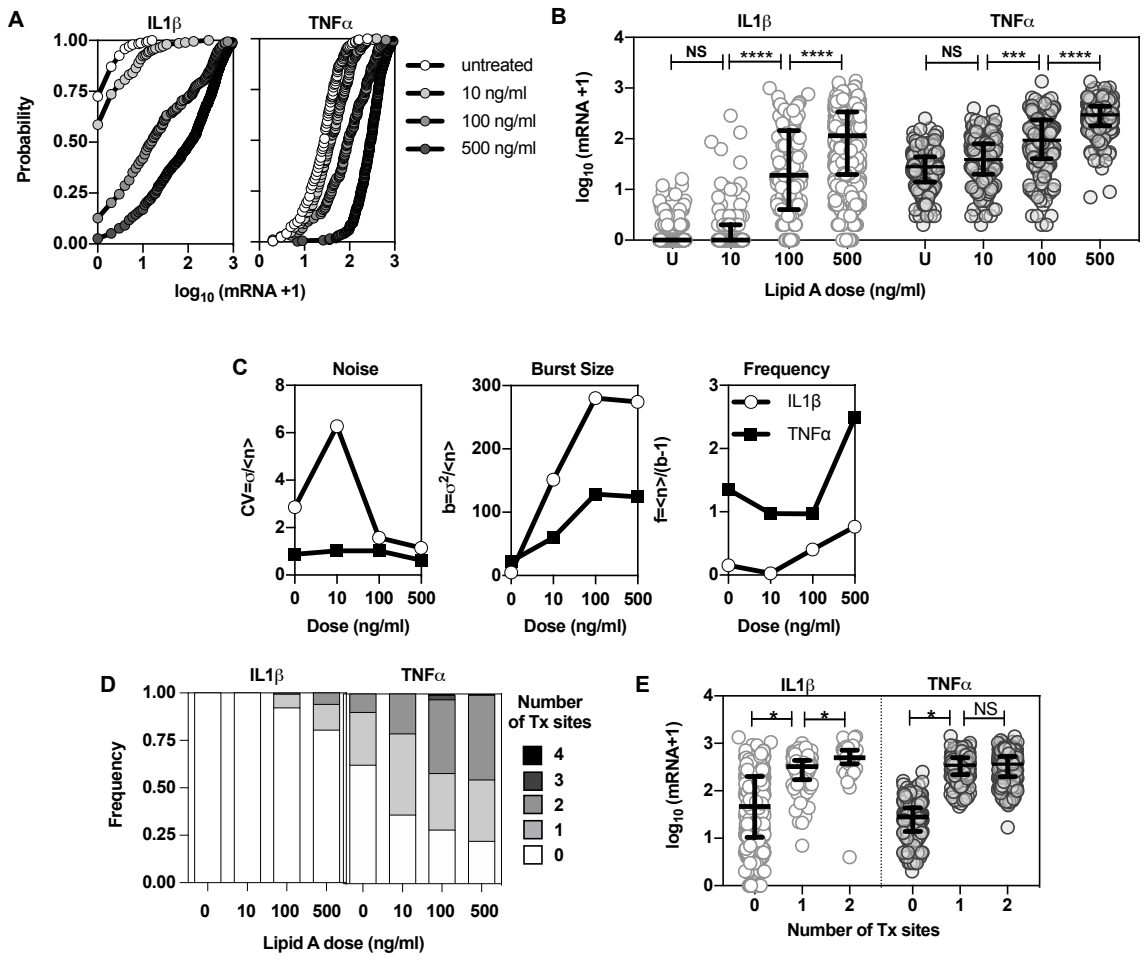


Figure S7. Dose-dependent regulation of *IL1β* and *TNFα* transcription in RAW 264.7 cells, related to Figure 2. A. Cumulative distribution function of mRNA counts in RAW 264.7 macrophages either untreated (U) or stimulated with 10, 100 and 500 ng/ml of lipid A for 3 h. 240, 208, 188 and 718 cells measured for *IL1β*, and 240, 208, 188 and 356 for *TNFα*, pooled across at least three smFISH experiments, respectively and expressed as $\log_{10}$(mRNA+1). **B**. Individual cell mRNA counts from A [with mean (and SD) per condition]. A nonparametric one-way ANOVA with Tukey's correction for multiple comparisons between groups summarised with ****- p-value <0.0001, ***- p-value <0.001, ** p-value <0.01, * p-value <0.05, NS- not significant. **C**. Coefficient of variation (*CV*), as well as moment estimators of burst size (*b*) and frequency (*f*) for mRNA count distributions from A. **D.** Distribution of transcription sites (Tx) in data from A. Shown is the fraction of cells with 0-4 transcription sites. **E.** mRNA abundance is correlated with the presence of transcription site. Shown are the *IL1β* and *TNFα* mRNA counts as a function of Tx number for cells stimulated with 500 ng/ml of lipid A (data from A). '*' denotes a statistical test (p-val<0.05) for a one-way ANOVA with Tukey's correction for multiple comparisons.

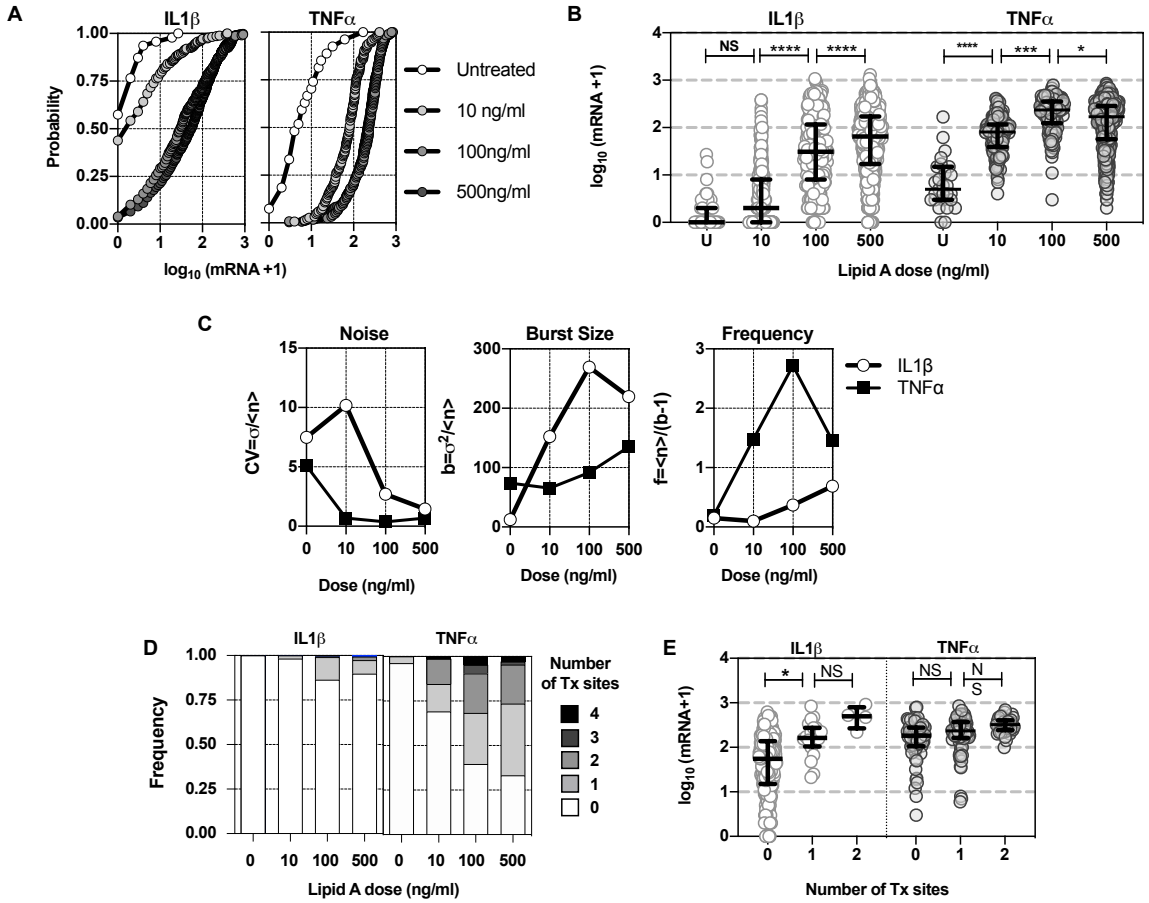


Figure S8. Dose-dependent regulation of *IL1β* and *TNFα* transcription in BMDMs, related to Figure 2. A. Cumulative distribution function of *IL1β* and *TNFα* mRNA counts in BMDMs either untreated (U, 0 ng/ml) or stimulated with 10, 100 and 500 ng/ml of lipid A for 3 h. 47, 276, 324 and 732 cells measured for *IL1β*, and 27, 149, 126 and 322 for *TNFα*, pooled across at least three smFISH experiments, and expressed as $\log_{10}$(mRNA+1). **B**. Individual cell mRNA counts from A. Shown are individual cell counts together with mean (and SD) per condition. Nonparametric Mann-Whitney U test for pairwise comparisons between groups summarised with ****- p-value <0.0001, ***- p-value <0.001, ** p-value <0.01, * p-value <0.05, NS- not significant. **C**. Characteristics of single cell mRNA expression. Shown is the coefficient of variation (*CV*) as well as moment estimators of burst size (*b*) and frequency (*f*) for mRNA count distributions from A. **D**. Distribution of transcription sites (Tx) in data from A. Shown is the fraction of cells with 0-4 transcription sites. **E**. *IL1β* and *TNFα* mRNA counts as a function of Tx number for cells stimulated with 500 ng/ml of lipid A (data from A). '*' denotes a statistical test (p-val<0.05) for one-way ANOVA with Tukey's correction for multiple comparisons.

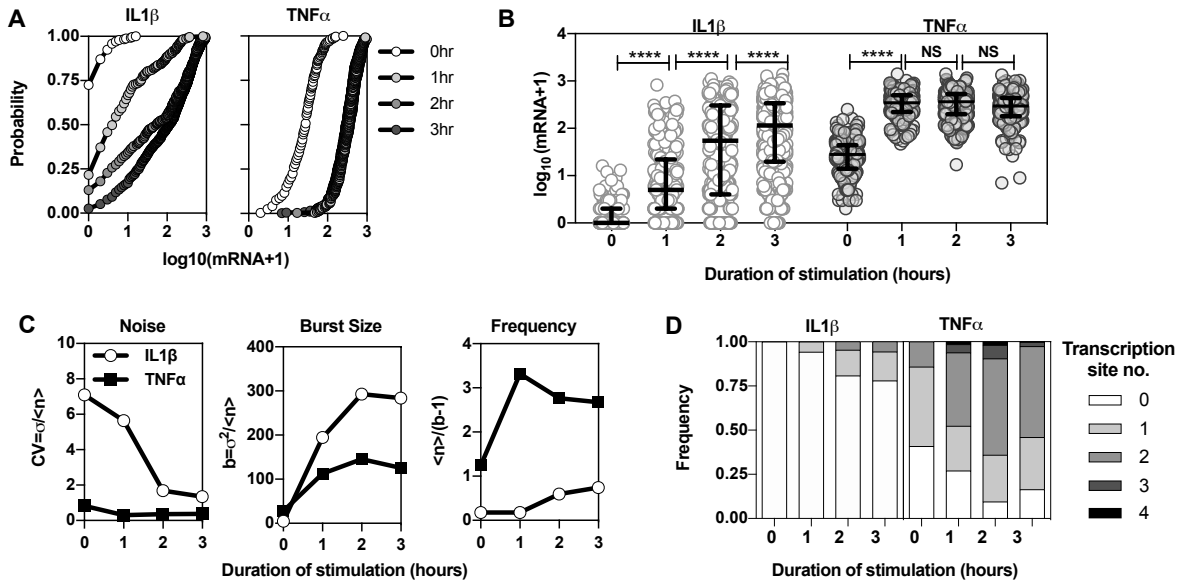


Figure S9. Variability of temporal *IL1β* and *TNFα* mRNA expression, related to Figure 2. A. Cumulative distribution function of mRNA counts in RAW 264.7 cells either untreated (0 h) or treated with 500 ng/ml lipid a for 1, 2 or 3 hours. Data pooled across at least three smFISH experiments and expressed as $\log_{10}$(mRNA+1). **B**. Individual cell mRNA counts in RAW 264.7 cells either untreated (0 h) or stimulated with 500 ng/ml of lipid A for 1, 2 and 3 h. 240, 253, 338 and 718 cells measured for IL1β, and 240, 253, 338 and 356 for TNFα, across at least three smFISH experiments, respectively. A nonparametric one-way ANOVA with Tukey's correction for multiple comparisons between groups summarised with ****- p-value <0.0001, ***- p-value <0.001, ** p-value <0.01, * p-value <0.05, NS- not significant. **C**. Characteristics of single cell mRNA expression. Shown is the coefficient of variation (*CV*) as well as moment estimators of burst size (*b*) and frequency (*f*) for mRNA count distributions from A. **D**. Distribution of transcription sites. Shown is the distribution of transcription sites in data from A. Shown is the fraction of cells with 0-4 transcription sites.

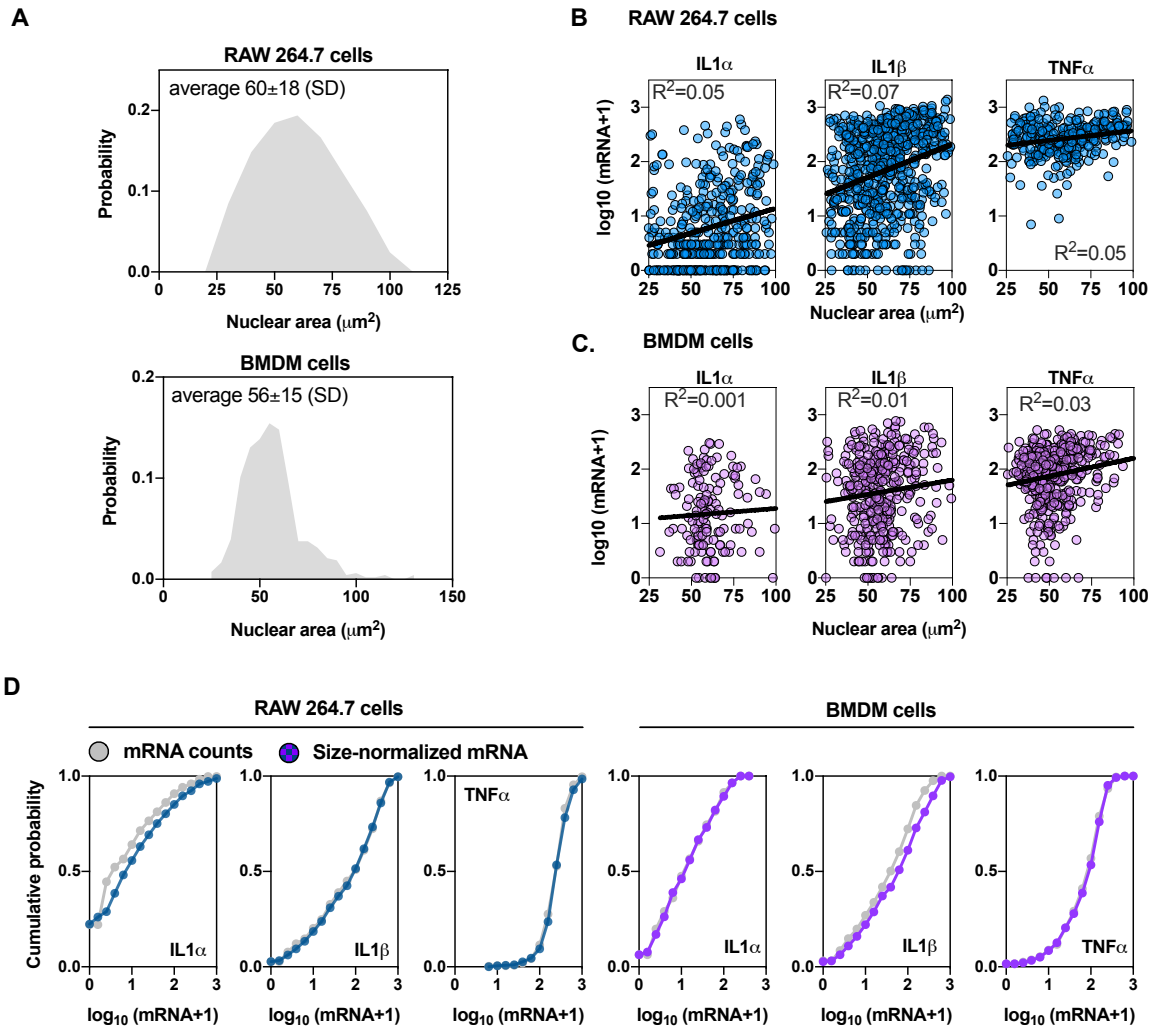


Figure S10. Effect of nuclear size on the smFISH count distribution, related to Figure 2. A. Smooth histogram of the nuclear area of RAW 264.7 and BMDM cells treated with 500 ng/ml lipid A for 3 h (data from Figs 1 and S8). Mean nuclear size and standard deviation (SD) displayed. **B and C.** Correlation between mRNA levels and the nuclear size. Shown are scatter plots across RAW 264.7 (B) and BMDM (C) cells stimulated with 500 ng/ml lipid A for 3 h (from data in Figs 1 and S8), with fitted regression line and fraction of variance explained by the nuclear size (correlation coefficient R^2). **D.** Cumulative probability distribution of the *IL1α*, *IL1β* and *TNFα* mRNA counts expressed as $\log_{10}$ (mRNA+1). Shown is the comparison between raw and cell size-normalized mRNA counts. Size-normalisation performed by scaling individual mRNA counts via the ratio of the average nuclear area in the population and nuclear size of particular cell.

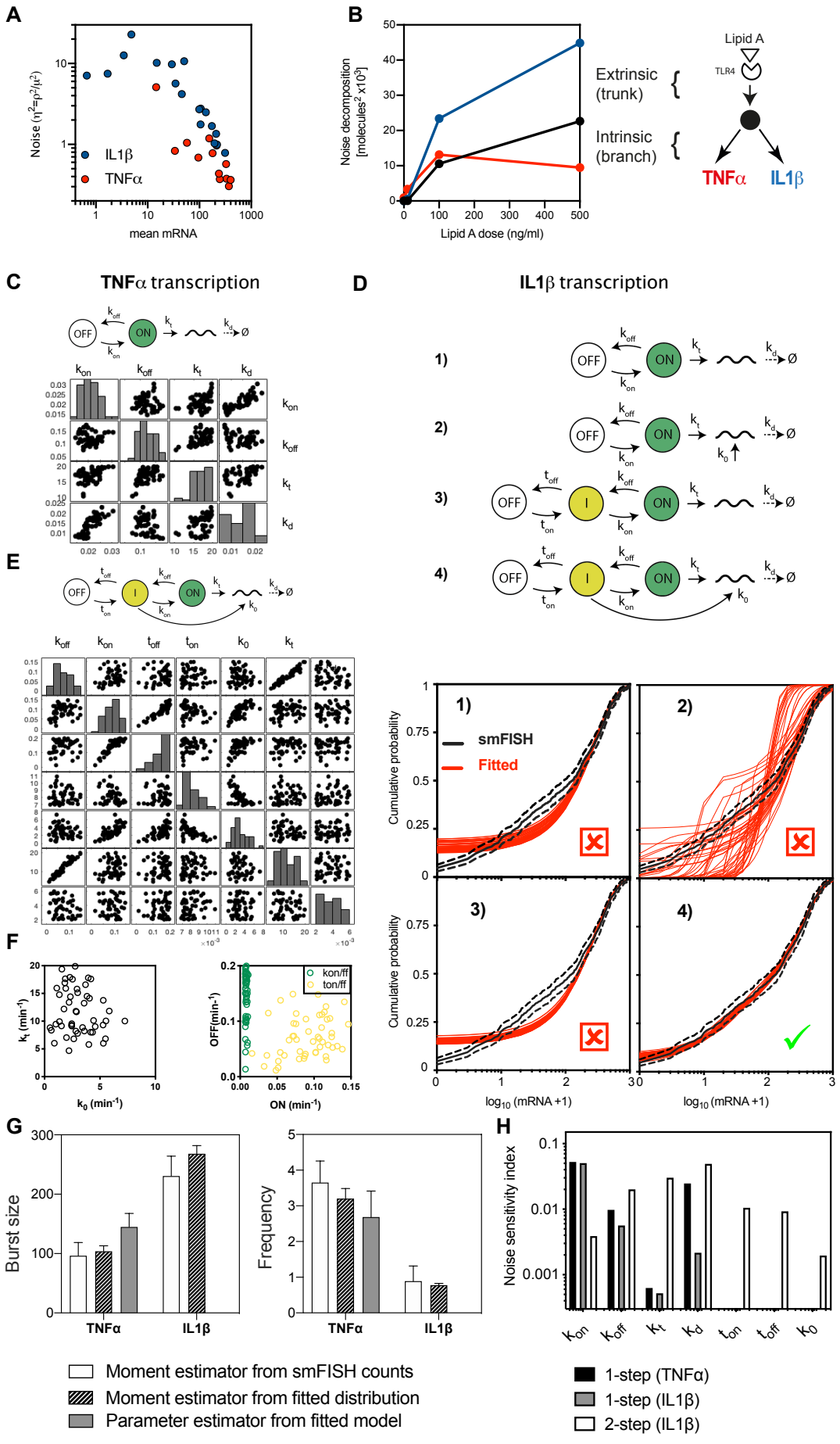

A
Noise ($\eta^2=\rho^2/\mu^2$)
IL1β
TNFα
mean mRNA
B
Noise decomposition [molecules^2 x10^3]
Lipid A dose (ng/ml)
Extrinsic (trunk)
Intrinsic (branch)
Lipid A
TLR4
TNFα
IL1β
C
TNFα transcription
OFF
ON
k_{on} k_{off} k_t k_d
D
IL1β transcription
1)
2)
3)
4)
I
t_{off} t_{on} k_0
E
k_{off} k_{on} t_{off} t_{on} k_0 k_t
F
k_t (min^{-1})
k_0 (min^{-1})
OFF(min^{-1})
ON (min^{-1})
kon/ff
ton/ff
Cumulative probability
smFISH
Fitted
$\log_{10}$ (mRNA +1)
G
Burst size
Frequency
TNFα
IL1β
H
Noise sensitivity index
k_{on} k_{off} k_t k_d t_{on} t_{off} k_0
Moment estimator from smFISH counts
Moment estimator from fitted distribution
Parameter estimator from fitted model
1-step (TNFα)
1-step (IL1β)
2-step (IL1β)

Figure S11. Mathematical model fits of *TNFα* and *IL1β*, related to Figure 2. A. Analysis of noise in the TNFα and IL1β smFISH counts. Show is $\eta^2=CV^2$ as a function mean mRNA expression (in $\log_{10}$ scale) combining all experimental conditions in this work (i.e., lipid A dose-response in RAW 264.7 and BMDM cells, time-course in RAW 264.7 and perturbations). **B**. Decomposition of noise in the *TNFα* and *IL1β* mRNA numbers; extrinsic noise (in black) versus intrinsic noise in *TNFα* (red) and *IL1β* (blue) levels calculated for the lipid A dose-response data (Fig. S8). Right: Schematic diagram of the noise decomposition. Trunk noise represents extrinsic variability between cells (potentially due to TLR signalling or generic transcriptional machinery), branch noise corresponds to gene specific intrinsic noise. **C**. Schematics of TNFα model shown alongside distribution of the fitted *TNFα* model parameters. Shown is a scatter plot matrix (with corresponding histograms) of individual 50 model fits from Fig. 2D (smFISH mRNA distribution at 3 h after 500 ng/ml lipid A stimulation in RAW 264.7 cells). **D**. *IL1β* transcription conforms to a two-step stochastic model. Top: Considered models of *IL1β* transcription: 1) one-step model with inducible transcription (k_t); 2) one-step model with a basal transcription (k_0); 3) basic two-step model; 4) two-step model. (Bottom) Shown is the comparison between measured and fitted *IL1β* mRNA distributions (3 h after 500 ng/ml lipid A treatment for the four different models in RAW 264.7 cells). In black: Kaplan-Meier estimator of measured CDF (with 95% confidence intervals), in red: a family of 50 models fitted to the data. **E**. Distribution of the fitted two-step model parameters from Fig. 2E (smFISH mRNA counts at 3 h after 500 ng/ml lipid A stimulation in RAW 264.7 cells). Shown is a scatter plot matrix (with histograms) of 50 individual model 4 fits (from B). **F**. *IL1β* model involves a combination of high and low transcription rates. Shown is the scatter plot of fitted transcription rates (k_t, k_0) for a family of 50 fitted models from Fig. 2E. (Right) Relationship between t_{on}/t_{off} rates (in yellow) and k_{on}/k_{off} rates (in green). **G**. Estimates of bursting characteristics for the fitted family of *TNFα* and *IL1β* models. Shown is the comparison between moment estimators for smFISH counts (from Fig. 2A), fitted distributions obtained with the model fits (from Fig. 2D and E) and estimates from based on fitted parameters values (for *TNFα).* **H.** Sensitivity analyses of different model structures**.** Shown is the local sensitivity analysis for the noise level (σ/μ) for the one-step *TNFα* and two-step *IL1β* models (as in Fig. 2D and E, respectively) as well as one-step model refitted to recapitulate mean and higher variance of *IL1β* expression (k_{on}=0.005 min^{-1}, k_{off}=0.05 min^{-1}, k_t=10 mRNA/min, k_d=0.003 min^{-1}). Sensitivity indexes calculated for 10% individual parameter changes.

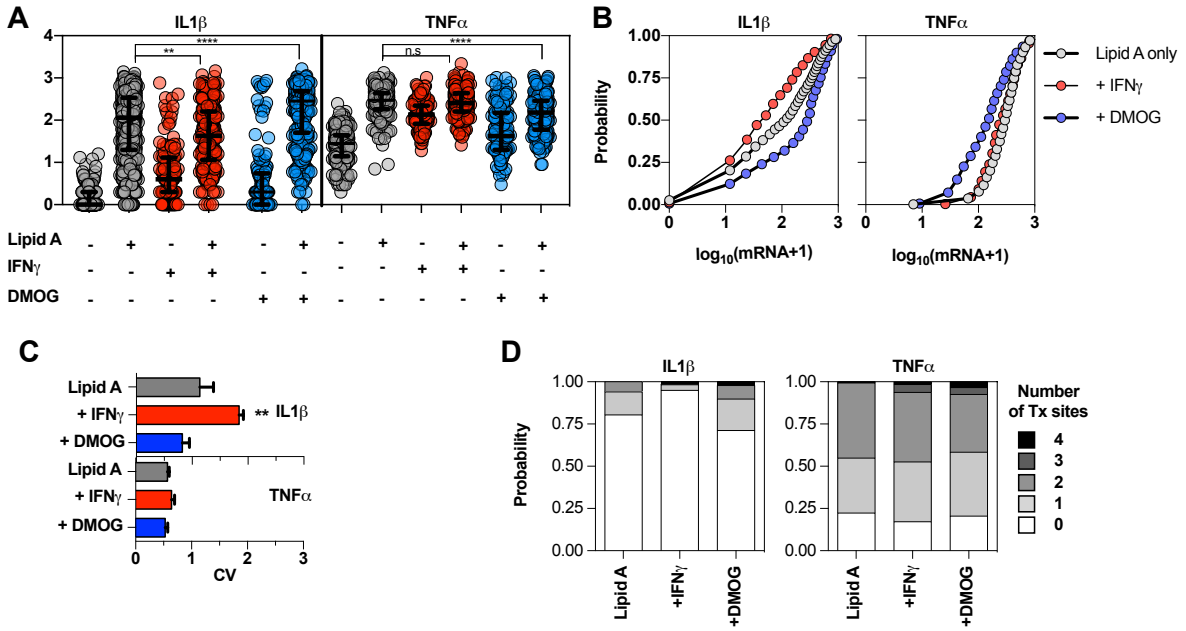


Figure S12. Perturbation of *TNFα* and *IL1β* gene expression by co-stimulation, related to Figure 3. A. smFISH analysis of *TNFα* and *IL1β* mRNA in response to co-stimulation with lipid A and DMOG or IFNγ. RAW 264.7 cells stimulated with 500 ng/ml of lipid A for 3 hours. For co-stimulation, cells were pre-treated with DMOG (0.5 mM) and IFNγ (100 ng/ml) for 24 hours before the treatment with lipid A. As a control, cells were stimulated with DMOG (0.5 mM) or IFNγ (100 ng/ml) for 27 hours. Nonparametric one-way ANOVA with Tukey's correction for multiple comparisons between lipid A-treated and co-stimulated groups summarised with ****- p-value <0.0001, **- p-value <0.01, ns- not significant. **B.** Cumulative distribution function of mRNA count data from A. **C.** Coefficient of variation (CV) for mRNA count data from A. '**' denotes significance <0.002 for a nonparametric one-way ANOVA with Tukey's correction for multiple comparisons. **D.** Distribution of transcription sites. Shown is the distribution of transcription sites in data A. Shown is the fraction of cells with 0-4 transcription sites.

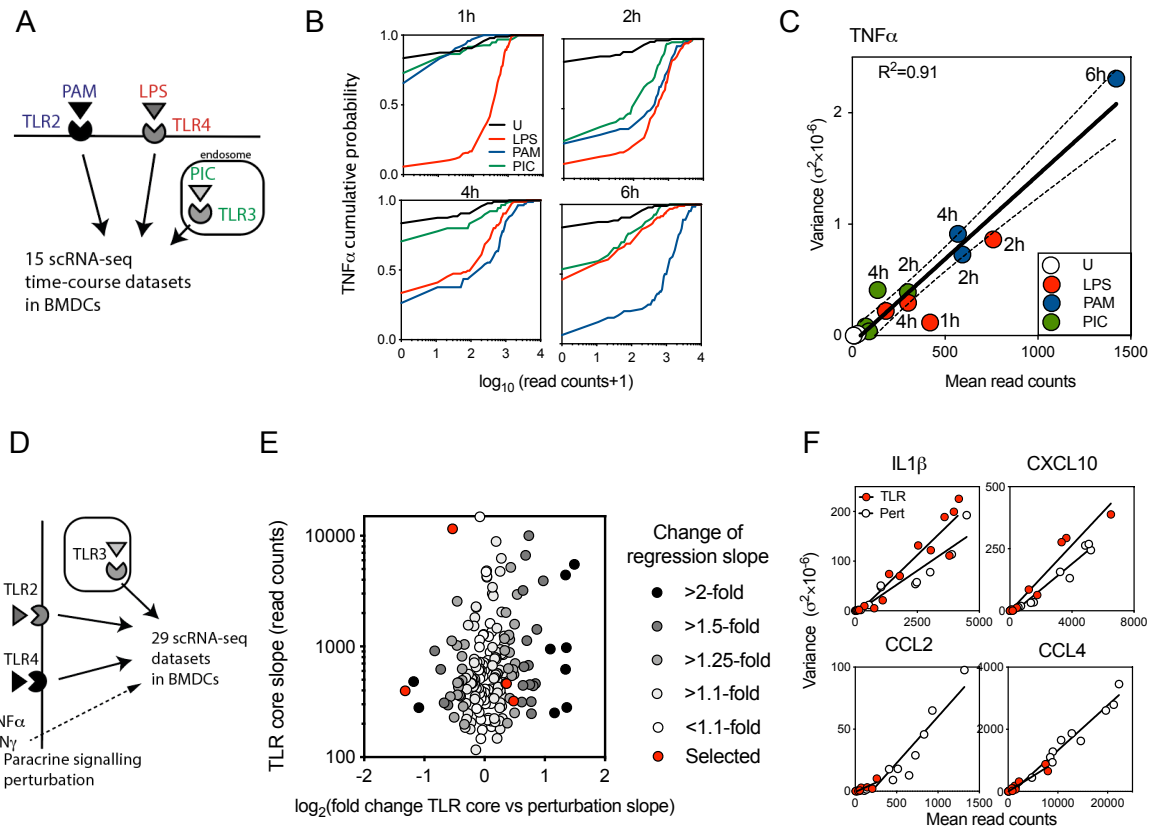


Figure S13. Analysis of TLR gene expression response heterogeneity in BMDCs, related to Figure 3. A. Schematic representation of the core TLR signalling network. **B.** Temporal TNFα expression patterns are treatment specific. Shown are cumulative distribution functions of TNFα read counts across different treatments [untreated (U), LPS, PAM and PIC, as indicated on the graph). Distributions estimated from the scRNA-seq data from (Shalek et al., 2014). **C**. The inferred linear regression trend (with 95% confidence intervals) for TNFα from [3] Differential TLR stimulation colour-coded as in A, highlighted are specific measurement times. Coefficient of determination depicted with R^2. **D.** Schematic representation of TLR paracrine signalling pathways. BMDC either stimulated with core TLR treatments, or perturbed using generic (e.g. Golgi inhibitors) and specific paracrine signalling modulators, e.g. using INFAR1, TNFR and STAT1 knockout cells. **E.** Effect of paracrine signalling perturbation on the fitted mean-variance regression trends. Shown is the fold-change of the regression slopes fitted for core TLR dataset vs. set of signalling perturbations (in $\log_2$) for 195 high confidence genes (defined by $R^2>0.75$ for both regression fits). Fold-change levels depicted in grey scale (as indicated on the graph). **F.** Fitted mean-variance relationships for selected genes from D. Data points corresponding to core TLR and perturbation indicated with red and open circles, respectively. Gene-specific regression slopes are statistically different (p-value <0.05) as assessed with a Student t-test.

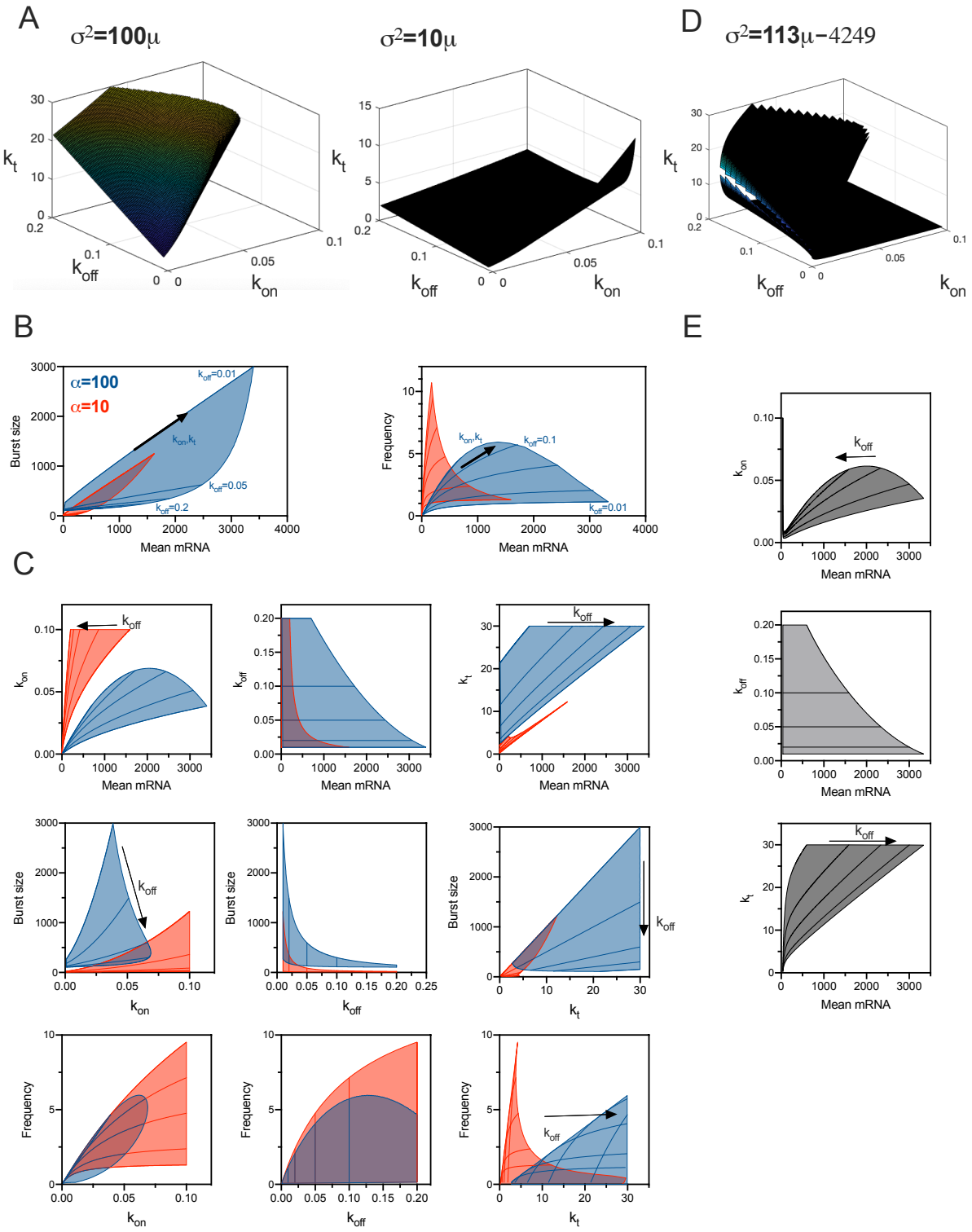


Figure S14. Linear mean-variance relationships constrain parameters of transcription, related to Fig. 4. A. Three-dimensional (k_{off}, k_{on}, k_t) parameter surface on which the $\sigma^2=\alpha\mu$ relationship holds. Calculation performed using Eq. (6) for biologically plausible set of gene activity switching rates, $k_{off}<0.2\ min^{-1}$ and $k_{on}<0.1\ min^{-1}$, while assuming $k_d=0.014\ min^{-1}$ (corresponding to fitted *TNFα* mRNA degradation rate) and $k_t<30\ min^{-1}$. Shown are relationships for $\alpha=100$ and $\alpha=10$, as highlighted on the graph. **B.** Burst size and burst

frequency as function of the mean mRNA response calculated for the kinetic parameter values given by surface from A. In different colours are the feasible parameter ranges (blue for $\alpha=100$, red for $\alpha=10$), thin lines correspond to $k_{off}=0.01,\ 0.02,\ 0.05,\ 0.1,\ 0.2\ min^{-1}$. **C**. Bursting characteristics as a function of kinetic parameter values given by surface from A. In different colours are the feasible parameter ranges (blue for $\alpha=100$, red for $\alpha=10$), thin lines correspond to $k_{off}=0.01,\ 0.02,\ 0.05,\ 0.1,\ 0.2\ min^{-1}$. **D.** Three-dimensional (k_{off}, k_{on}, k_t) parameter surface on which the $\sigma^2=\alpha\mu-\alpha_0$ relationship holds. Calculation performed using Eq. (7) for $\alpha=113$ and $\alpha_0=4249$ (corresponding to the fitted $TNF\alpha$ relationship) for biologically plausible set of gene activity switching rates, $k_{off}<0.2\ min^{-1}$ and $k_{on}<0.1\ min^{-1}$, while assuming $k_d=0.014\ min^{-1}$ and $k_t<30\ min^{-1}$. Highlighted lines correspond to $k_{off}=0.01,\ 0.02,\ 0.05,\ 0.1,\ 0.2\ min^{-1}$. **E**. Kinetic parameter values as a function of mean mRNA expression given by surface in D.

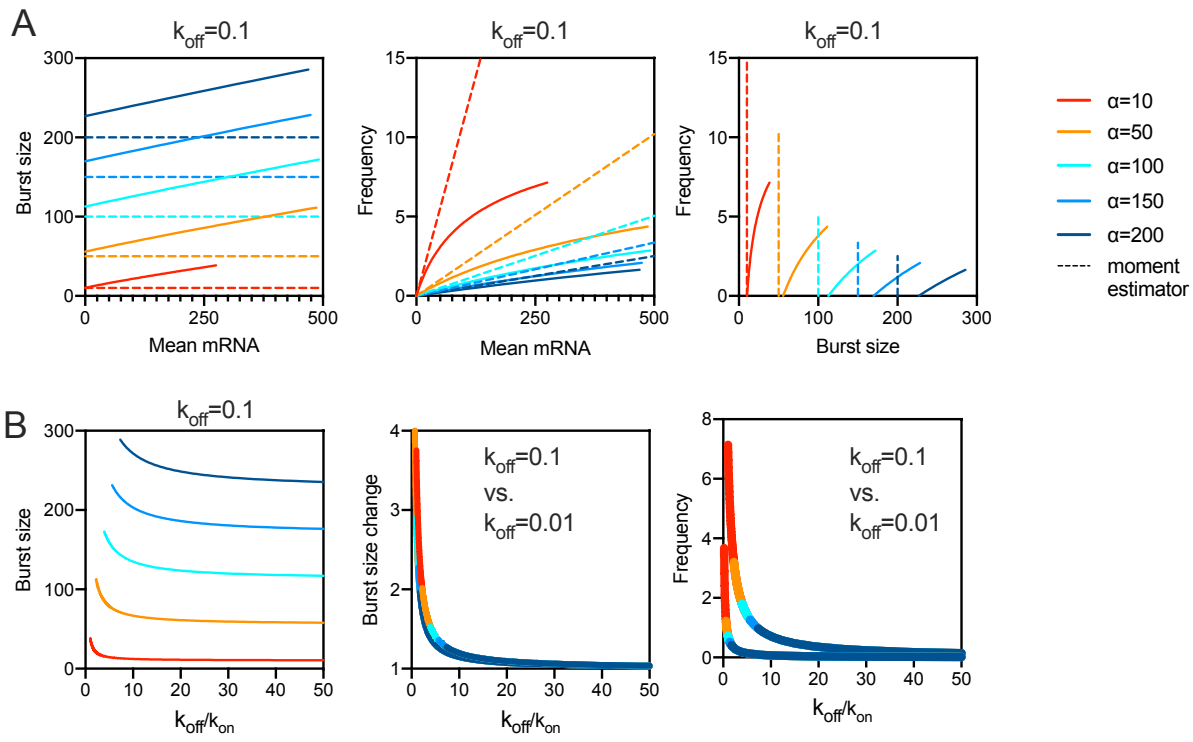
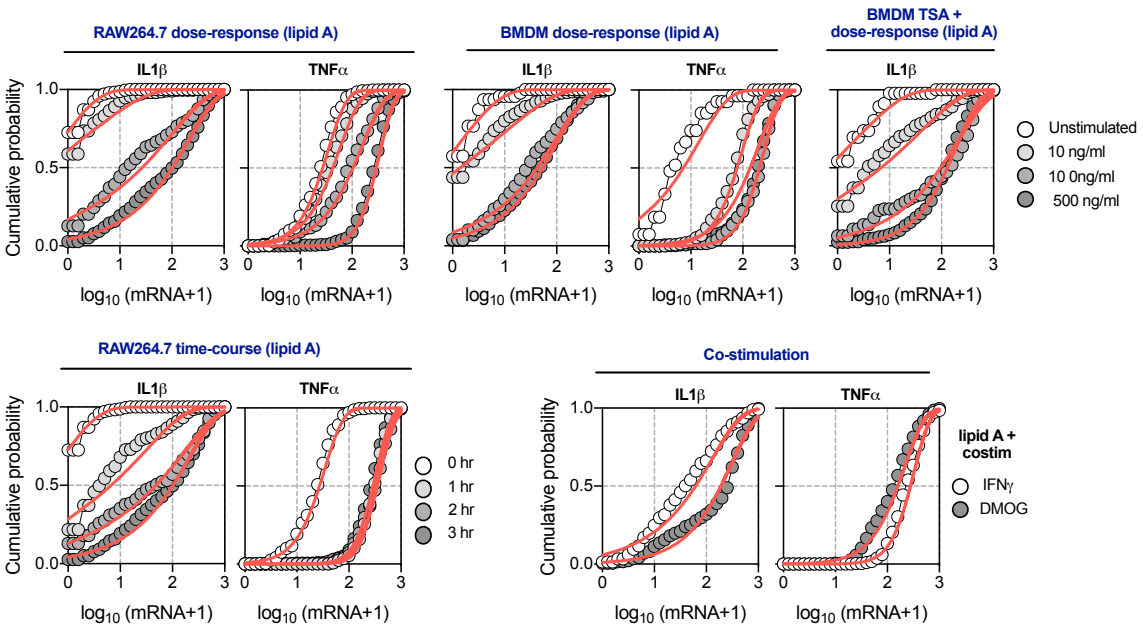


Figure S15. Reciprocal relation between the burst size and frequency, related to Fig. 4. **A.** Changes of burst size and frequency across the underlying range of mean mRNA expression. Calculation performed using Eq. (6) for biologically plausible set of gene activity switching rates $k_{on}<0.1\ min^{-1}$, while assuming $k_d=0.014\ min^{-1}$ (corresponding to fitted *TNFα* mRNA degradation rate), $k_t<30\ min^{-1}$ and μ<500. Fife putative genes are considered, each with a different regression slope (α, in different colour lines). Shown are relative frequency and burst size changes per α, corresponding to $k_{off}=0.1\ min^{-1}$. In broken lines are the moment estimators (i.e., '*bursty*' regime). B**.** Analysis of absolute burst size and frequency from B as a function of k_{off}/k_{on} ratio. Colour coding as in A. Calculations performed for $k_{off}=0.1$ and $0.01\ min^{-1}$.

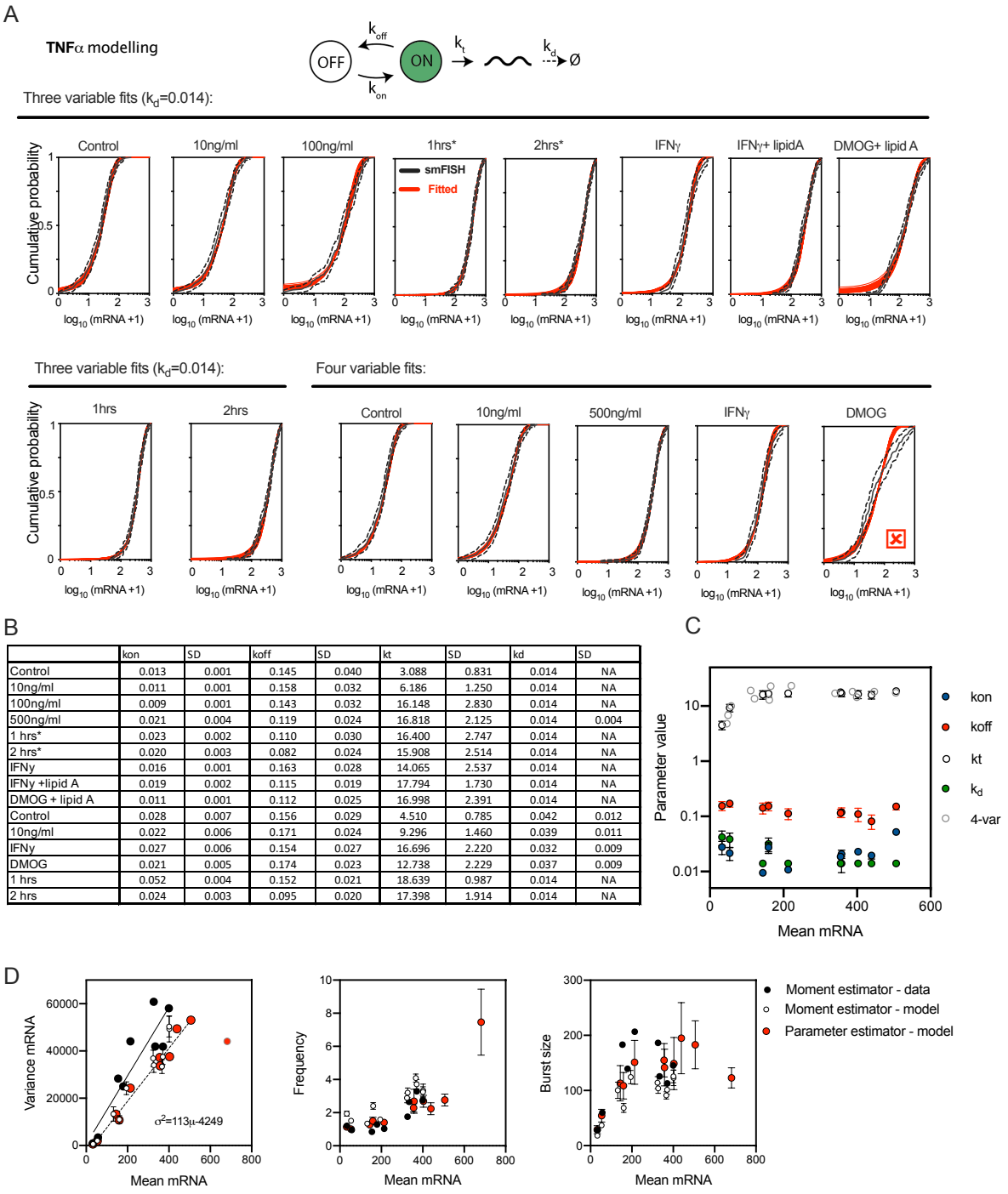
A



B

		TNFα			IL1β		
Cell Line	Condition	r	p	NB p-val	r	p	NB p-val
RAW264.7	Unstimulated	1.56	0.04	0.25	0.23	0.26	N/A
	Lipid A (3 hrs; 10 ng/ml)	1.18	0.02	0.09	0.14	0.03	N/A
	Lipid A (3 hrs; 100 ng/ml)	0.93	0.01	0.17	0.30	0.00	0.014
	Lipid A (3 hrs; 500 ng/ml)	2.60	0.01	0.29	0.54	0.00	<0.001
	Lipid A (1 hr; 500 ng/ml)	3.25	0.01	0.33	0.26	0.01	0.024
	Lipid A (2 hrs; 500 ng/ml)	2.53	0.01	0.02	0.32	0.00	<0.001
	IFNγ	1.81	0.01	0.12	0.24	0.01	N/A
	IFNγ + Lipid A	2.19	0.01	0.02	0.51	0.00	0.014
	DMOG	0.67	0.01	0.003	0.10	0.00	N/A
	DMOG + Lipid A	1.24	0.01	0.09	0.76	0.00	<0.001
BMDM	Unstimulated	0.52	0.03	N/A	0.25	0.13	N/A
	Lipid A (3 hrs; 10 ng/ml)	1.67	0.02	0.24	0.17	0.01	0.069
	Lipid A (3 hrs; 100 ng/ml)	2.04	0.01	0.03	0.46	0.00	0.003
	Lipid A (3 hrs; 500 ng/ml)	1.80	0.01	0.051	0.58	0.01	0.586
	TSA	N/A	N/A	N/A	0.20	0.05	N/A
	TSA + Lipid A (3 hrs; 10 ng/ml)	N/A	N/A	N/A	0.23	0.00	0.002
	TSA + Lipid A (3 hrs; 100 ng/ml)	N/A	N/A	N/A	0.48	0.00	0.027
	TSA + Lipid A (3 hrs; 500 ng/ml)	N/A	N/A	N/A	0.88	0.00	0.196

Figure S16. Negative binomial fits of the measured *IL1β* and *TNFα* mRNA distributions, related to Fig 4. A. Comparison between negative binomial fit (in red) and measured mRNA distributions (depicted with different colour dots) across all smFISH datasets. **B.** Fitted negative binomial parameters (*r* and *p*) across different conditions. P-values denote result for a chi-squared test for the smFISH count distribution following negative binomial with respective parameters. N/A denotes cases when the test cannot be performed (due to low count levels) or measurement is not obtained.

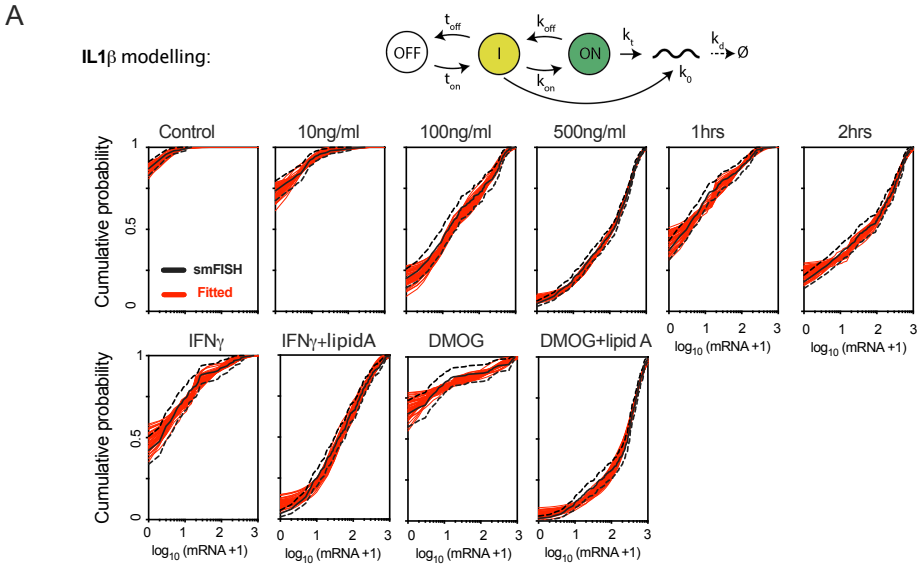


	kon	SD	koff	SD	kt	SD	kd	SD
Control	0.013	0.001	0.145	0.040	3.088	0.831	0.014	NA
10ng/ml	0.011	0.001	0.158	0.032	6.186	1.250	0.014	NA
100ng/ml	0.009	0.001	0.143	0.032	16.148	2.830	0.014	NA
500ng/ml	0.021	0.004	0.119	0.024	16.818	2.125	0.014	0.004
1 hrs*	0.023	0.002	0.110	0.030	16.400	2.747	0.014	NA
2 hrs*	0.020	0.003	0.082	0.024	15.908	2.514	0.014	NA
IFNy	0.016	0.001	0.163	0.028	14.065	2.537	0.014	NA
IFNy +lipid A	0.019	0.002	0.115	0.019	17.794	1.730	0.014	NA
DMOG + lipid A	0.011	0.001	0.112	0.025	16.998	2.391	0.014	NA
Control	0.028	0.007	0.156	0.029	4.510	0.785	0.042	0.012
10ng/ml	0.022	0.006	0.171	0.024	9.296	1.460	0.039	0.011
IFNy	0.027	0.006	0.154	0.027	16.696	2.220	0.032	0.009
DMOG	0.021	0.005	0.174	0.023	12.738	2.229	0.037	0.009
1 hrs	0.052	0.004	0.152	0.021	18.639	0.987	0.014	NA
2 hrs	0.024	0.003	0.095	0.020	17.398	1.914	0.014	NA

Figure S17. Model analysis of *TNFα* distributions, related to Figure 4. A. The comparison between measured and fitted *TNFα* mRNA distributions across all experimental conditions in RAW 264.7 cells (including dose- and time-response as well as perturbations). In black: Kaplan-Meier estimator of measured CDF (with 95% confidence intervals), in red: a family of 50 two-step models fitted to the data. Subset of conditions fitted assuming fixed k_d (three variable fits). * denotes data collected at 1 or 2 h, but fitted at 3 h after stimulation. **B**. Summary of fitted parameter values (mean and standard deviation, SD) from A. **C.** Parameter values as

a function of the fitted mean mRNA expression per condition (using theoretical steady-state levels). **D**. Estimates of variability and bursting characteristics for the fitted *TNF*α models. Shown is the comparison between moments and moments estimators for smFISH counts (in black) and mRNA distributions obtained with the model fits (with open circles). Also shown are estimators based on fitted kinetic parameter rates (in red). Characteristics represented as function of the corresponding mean values. Regression lines fitted to mean-variance data for smFISH counts (solid line) and steady-state mean/variance calculated for fitted model parameters (broken line, with the corresponding equation displayed).

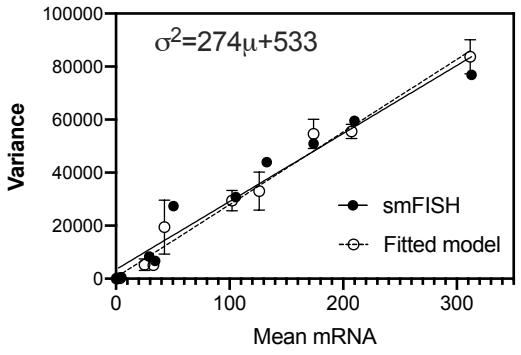
A



B

	ton	SD	toff	SD	kon	SD	koff	SD	k0	SD	kt	SD	kd	SD
Control	0.002	0.002	0.153	0.053	0.053	0.061	0.126	0.074	0.262	0.208	1.021	2.397	0.0037	NA
10ng/ml	0.002	0.003	0.146	0.052	0.029	0.027	0.124	0.060	0.664	0.492	5.833	5.731	0.0037	NA
100ng/ml	0.006	0.001	0.160	0.035	0.060	0.022	0.073	0.041	2.154	1.059	13.492	5.156	0.0037	NA
500ng/ml	0.008	0.001	0.144	0.045	0.091	0.031	0.072	0.036	3.054	1.513	11.683	4.108	0.0038	0.001
1hrs	0.010	0.003	0.154	0.038	0.062	0.024	0.114	0.048	1.202	0.595	11.864	4.210	0.0037	NA
2 hrs	0.007	0.002	0.136	0.038	0.130	0.038	0.081	0.050	3.492	2.308	13.148	4.734	0.0037	NA
IFNγ	0.003	0.001	0.160	0.039	0.035	0.019	0.083	0.052	1.235	0.659	11.150	5.055	0.0037	NA
IFNγ+ lipid A	0.008	0.001	0.147	0.045	0.044	0.019	0.075	0.038	2.710	1.359	13.966	4.460	0.0037	NA
DMOG	0.002	0.001	0.152	0.038	0.071	0.036	0.045	0.034	1.781	1.538	14.240	4.763	0.0037	NA
DMOG + lipid A	0.011	0.004	0.141	0.040	0.125	0.040	0.090	0.039	3.316	2.438	12.704	3.870	0.0037	NA

C



D

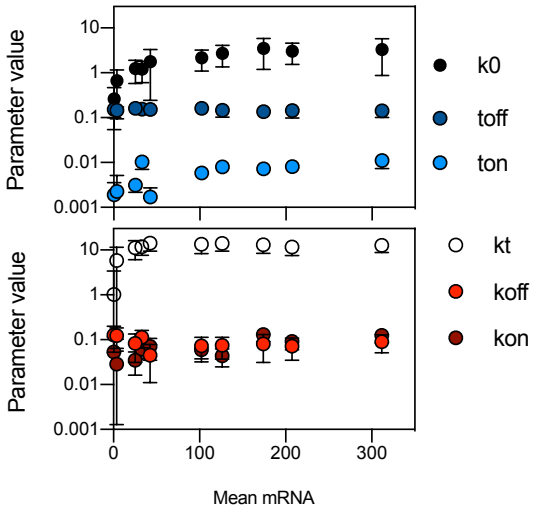


Figure S18. Model analysis of *IL1β* distributions, related to Figure 4. A. The comparison between measured and fitted *IL1β* mRNA distributions across all experimental conditions in RAW 264.7 cells (including does and time-response as well as perturbation). In black: Kaplan-Meier estimator of measured CDF (with 95% confidence intervals), in red: a family of 50 two-step models fitted to the data. **B**. Summary of fitted parameter values (mean and standard deviation, SD) from A. Models were fitted assuming fixed k_d (except of 500 ng/ml lipid A stimulation). **C**. Mean-variance relationship for the fitted family of *IL1β* models. Shown is the

comparison between moments for smFISH counts and mRNA distributions obtained with the model fits (with SDs). Regression lines fitted to smFISH counts (full circles) and mean/variance calculated for fitted model distributions (open circles with the corresponding equation displayed). **D.** Parameter values as a function of the fitted mean mRNA expression per condition.

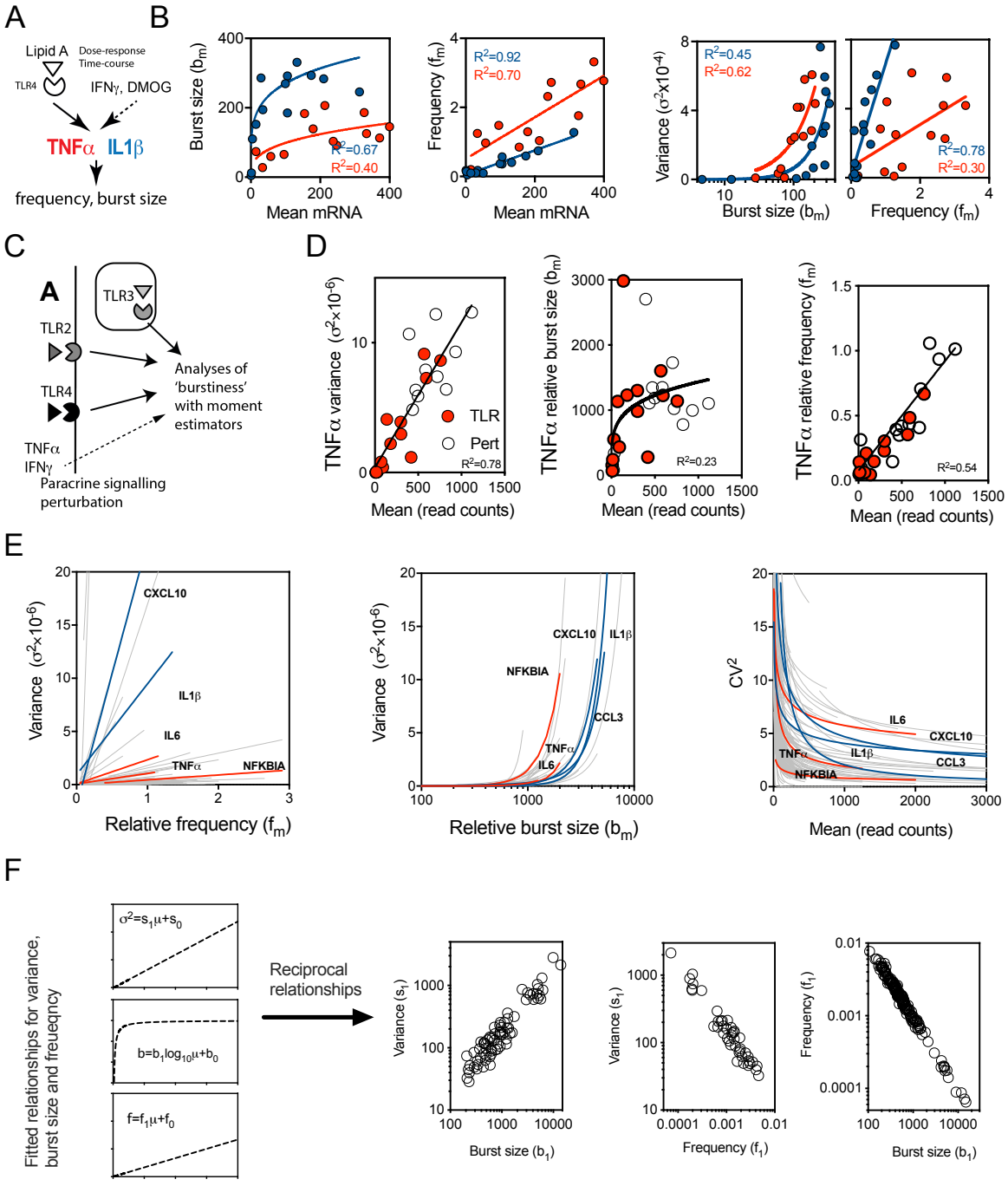


Figure S19. Analysis of transcriptional bursting using moment estimators, related to Figure 4. A. Schematic representation of immune-modulating pathways used to assay TNFα and IL1β mRNA expression. **B**. Moment estimators of transcriptional bursting in the smFISH dataset. Shown are individual data points and fitted relationships (in blue and red for *IL1β* and *TNFα*, respectively) for smFISH data (dose-response, time-course, as well as IFNγ and DMOG perturbation for BMDMs and RAW 264.7 macrophages). Power functions used to fit burst size (otherwise linear regression was applied). Coefficient of determination depicted with R^2 (colour coded for the respective gene). **C**. Schematic representation of the TLR and paracrine

signalling pathways. **D**. Transcriptional bursting of *TNFα* in the 26 scRNA-seq datasets in BMDCs from [3]. Shown is the fitted relationships combining core TLR (in red circles) and perturbation (in open circles) datasets using relative moment estimators (based on read counts). Coefficient of determination depicted with R^2. **E**. Fitted relationships for the variance and noise of TLR-induced genes, for the combined core TLR and perturbation dataset from [3]. Left: Relative frequency-variance relationship for the 112 genes (defined by Spearman correlation coefficient $R^2>0.5$) obtained using robust linear regression fit (with intercept). Middle: Relative burst size-variance relationship for the 189 genes (defined by Spearman correlation coefficient $R^2>0.68$) obtained using robust power series fit ($log_{10}(\sigma^2)=p1+p2\cdot\mu^{p3}$). Right: Mean-noise ($CV^2$) relationship for the 180 genes (defined by Spearman correlation coefficient $R^2>0.70$) obtained using robust curve fitting ($CV^2=p1\cdot\mu^{p2}$). **F.** Reciprocal relationship between variance, burst size and frequency in the fitted bursting characteristics (from Fig. 4F). Shown are pairwise scatter plots between fitted regression coefficients (as depicted on the left panel).

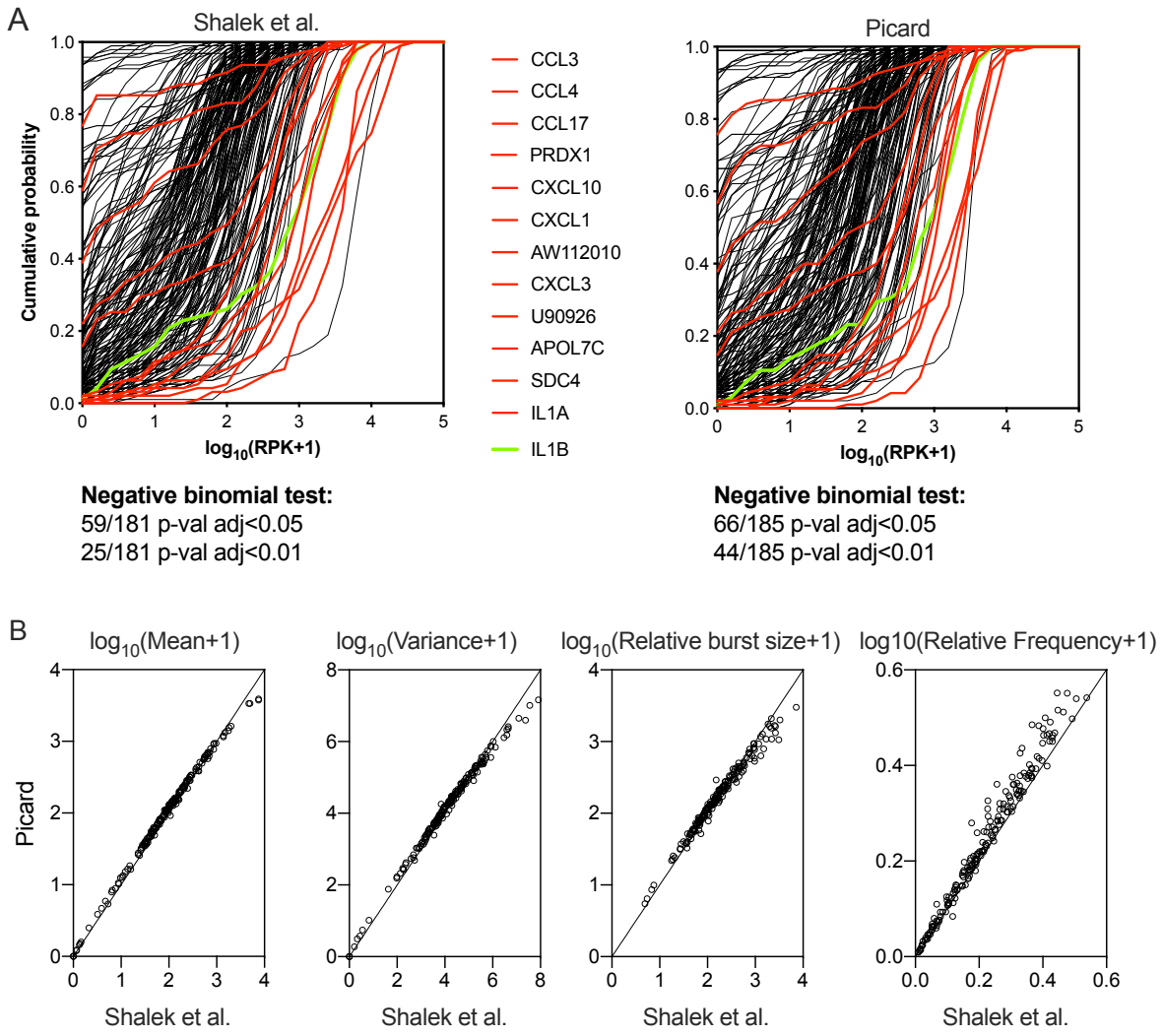


Figure S20. Validation of scRNA-seq measurements, related to Figure 4. A. Comparison between single cell distributions from (Shalek et al., 2014). LPS stimulation at 4h dataset and dataset obtained after re-mapping with Picard Tools (http://broadinstitute.github.io/picard/). Shown are cumulative distributions for 204 LPS-regulated genes from Fig. 3D, highlighted in red are high variability genes, in green *IL1β*. Shown are also fractions of genes (after removing low abundant genes), distribution of which does not fit negative binomial (as highlighted with chi-squared goodness-of-fit test p-values adjusted with Benjamini-Hochberg procedure for false discovery rate). **B.** Comparison between bursting characteristics for 204 LPS-regulated genes from A, in single cell distributions from (Shalek et al., 2014) and dataset obtained after re-mapping with Picard Tools. Characteristics presented in $\log_{10}$ scale, in black identity line.

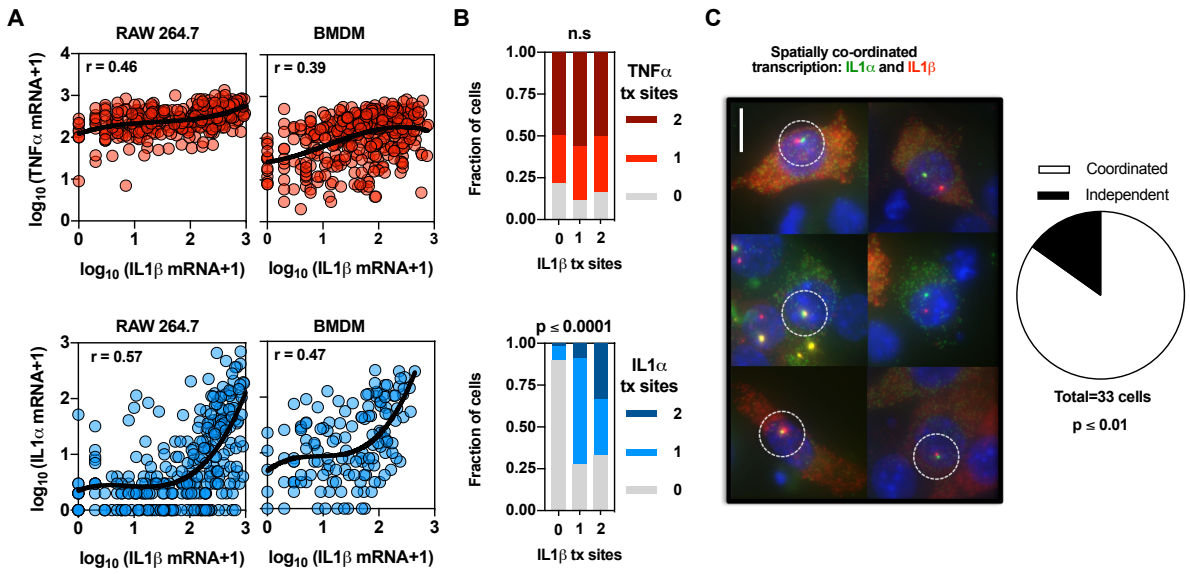


Figure S21. ***IL1α*** **and** ***IL1β*** **mRNA expression is coordinated in single cells at common transcription sites, related to Figure 5. A**. Correlation for *IL1β* vs. *TNFα* (top, in red) and *IL1β* vs. *IL1α* (bottom, in blue) mRNA counts. Shown are scatter plots of individual cell data (together with a cubic spline fit and a Spearman rank correlation *r*). RAW 264.7 and BMDM cells stimulated with 500 ng/ml of lipid A for 3 h. **B**. Correlation between transcription sites. Shown are the proportions of the co-activated transcription start sites for *IL1β* vs. *TNFα* (top) and *IL1β* vs. *IL1α* (bottom) for data in B. Shown also is a p-value for a chi-square test for the independence of Tx site occurrences between genes. **C**. Spatial analysis of *IL1α* and *IL1β* transcription. Deconvolved wide-field microscopy images of single cell smFISH mRNA counts for *IL1β* vs. *IL1α*. Circles indicate spatial coordination of transcription for data from A. Pie chart shows the proportion of spatially coordinated transcriptional sites for all cells in A exhibiting a common *IL1β* and *IL1α* Tx site (with a p-value for a chi-square test for independence).

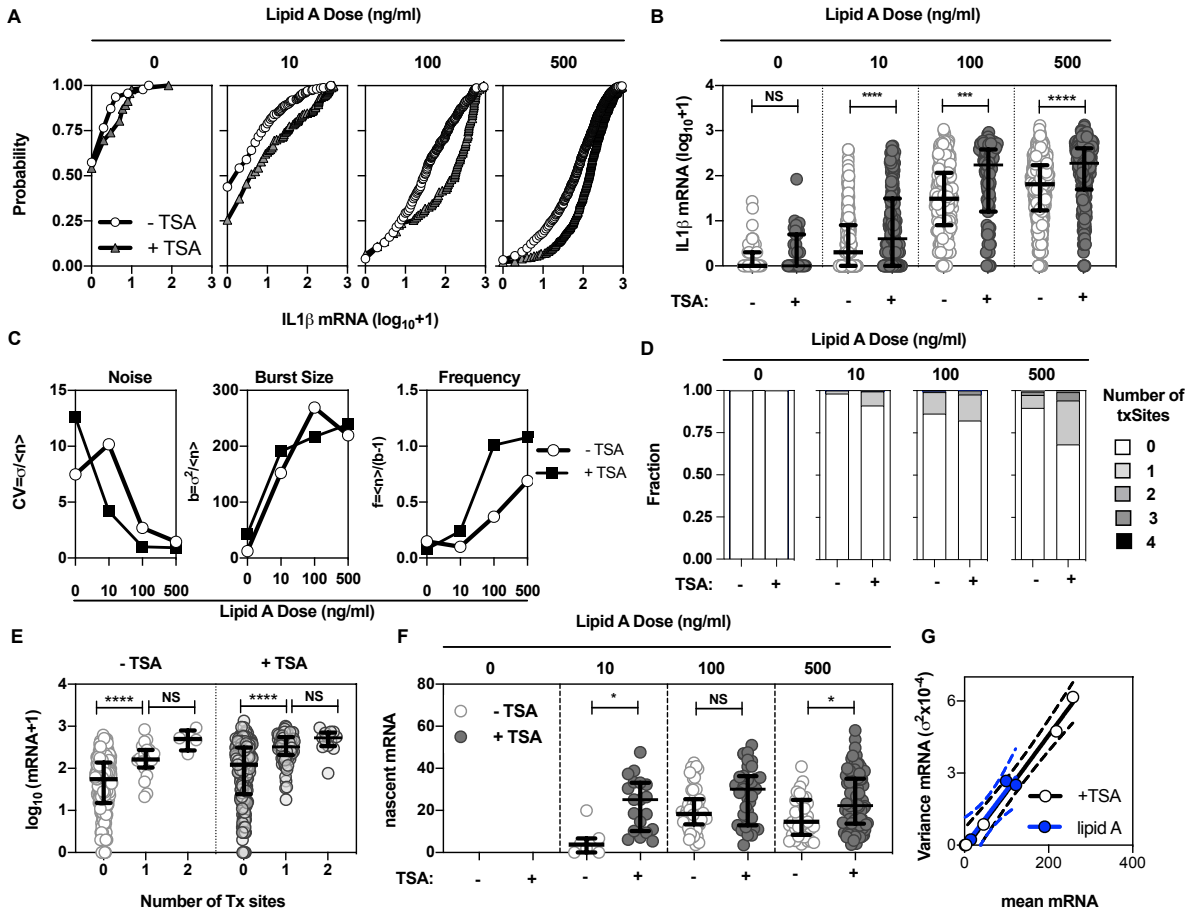


Figure S22. TSA modulation of *IL1β* mRNA distribution across lipid A dose-response, related to Figure 5. A. Shown are cumulative distribution plots of *IL1β* mRNA in BMDM cells pre-treated with TSA for 1 h (+TSA) or control group (-TSA) before stimulation with different lipid A doses for an additional 3 h. **B**. Individual cell mRNA counts data from A. Shown are data for 47, 276, 324 and 732 control (-TSA) as well as 46, 204, 110 and 305 (+TSA) for 0, 10, 100 and 500 ng/ml of lipid A, respectively. Results of the pairwise Mann-Whiney U tests summarised with ****- p-value <0.0001, ***- p-value <0.001, NS- not significant. **C**. Noise, burst size and frequency characteristics derived from data in B. **D**. Distribution of transcription sites observed in smFISH images from A. **E**. *IL1β* mRNA counts as a function of Tx site number for cells for 500 ng/ml lipid A dose. '*' denotes a statistical test (p-val<0.05) for one-way ANOVA with Tukey's correction for multiple comparisons. **F**. Quantification of nascent mRNA across lipid A dose-response from data in D. Shown is comparison across 0, 7, 52 and 35 (-TSA) as well as 0, 19, 35 and 114 (+TSA) transcription sites for 0, 10, 100 and 500 ng/ml of lipid A, respectively. Results of the pairwise nonparametric Mann-Whiney U tests summarised with ****- p-value <0.0001 and ***- p-value <0.001. **G.** Mean-variance relationship obtained for smFISH data for *IL1β* from A. Shown are the fitted regression lines

(with 95% confidence intervals in broken lines) together with individual data points. In blue cells pre-treated with TSA, in black lipid A treatment alone.

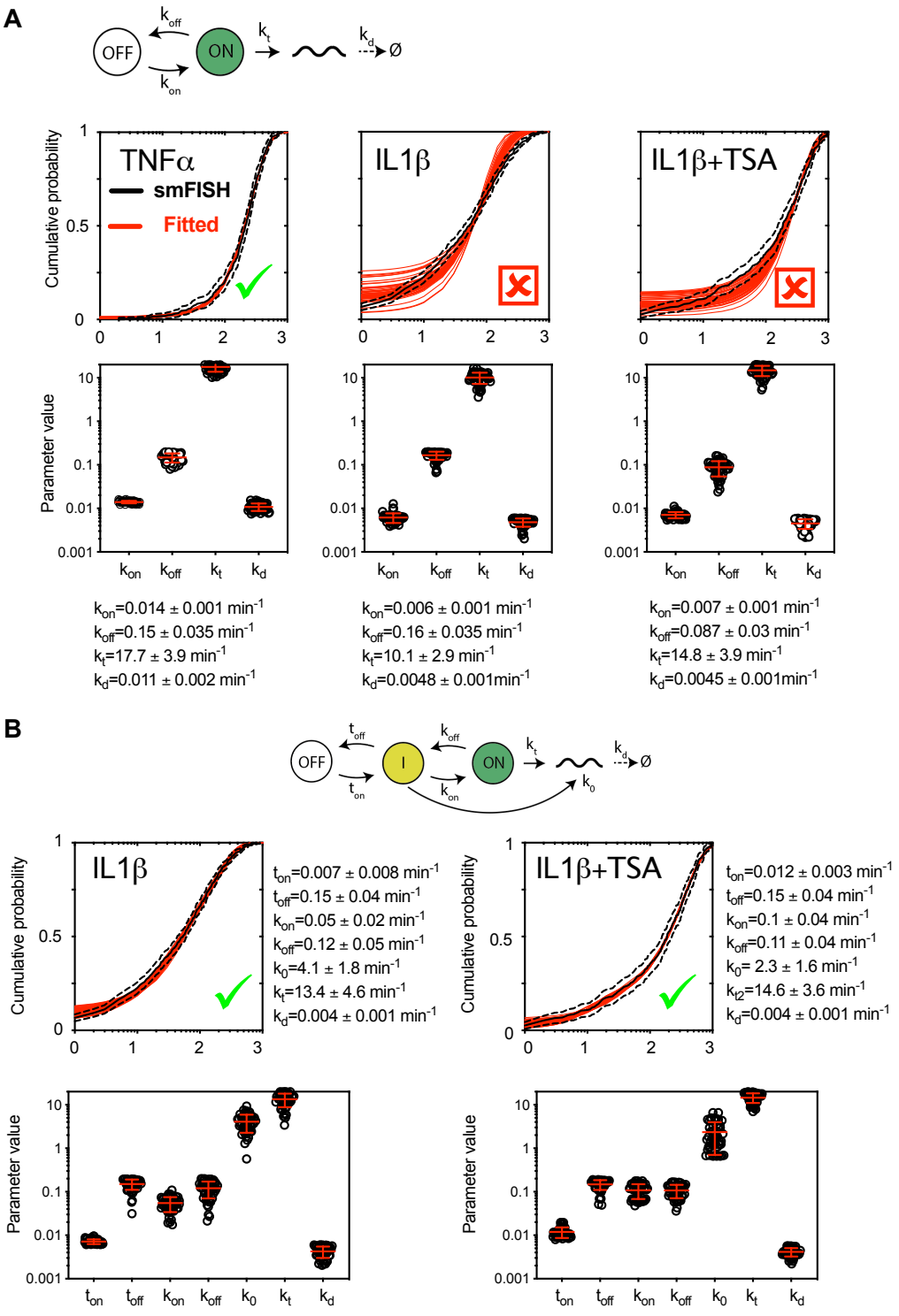


Figure S23. Characteristics of the fitted family of models for BMDM data, related to Figure 5. Distribution of the model parameters fitted to smFISH mRNA data at 3 h after 500 ng/ml lipid A stimulation in BMDM cells (in combination with TSA treatment, as in Fig. 5). Shown are parameters of 50 individual fits for one-step (**A**) and two-step (**B**) models summarised using a box plot (with mean ± SD in red). Fit quality summarised with green ticks and red crosses.

Chapter 3 Global characterisation of transcriptional variability in the TLR system and through evolution

In this chapter, I verify the linear trends and their underlying transcriptional bursting characteristics on a large-scale in the TLR system. Using a mathematical analysis approach based on inference of mean-variance linear relationships on an available scRNA-seq dataset, I establish global regulatory mechanisms of burst size and frequency to maintain controlled cell-to-cell variability in immune genes expression. Through stochastic modelling I show that increased level of variability is associated with greater complexity in transcriptional regulatory networks, and in this case the commonly used telegraph model of gene expression is not enough to capture the highly noisy transcriptional output. Next, I move onto exploring the transcriptional variability trait in the TLR system through evolution across different species; is transcriptional variability constrained in the same manner across different species? Is the level of variability of a gene conserved through evolution? Do transcriptional bursting characteristics patterns change during evolution? These are some of the questions answered in this paper.

3.1 Journal paper: Variability of the innate immune response is globally constrained by transcriptional bursting

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 27 June 2023
DOI 10.3389/fmolb.2023.1176107

OPEN ACCESS

EDITED BY
Guido Tiana,
University of Milan, Italy

REVIEWED BY
Attila Becskei,
University of Basel, Switzerland
Pavel Kos,
Friedrich Miescher Institute for Biomedical Research (FMI), Switzerland

*CORRESPONDENCE
Pawel Paszek,
pawel.paszek@manchester.ac.uk

RECEIVED 25 April 2023
ACCEPTED 15 June 2023
PUBLISHED 27 June 2023

CITATION
Alachkar N, Norton D, Wolkensdorfer Z, Muldoon M and Paszek P (2023), Variability of the innate immune response is globally constrained by transcriptional bursting.
Front. Mol. Biosci. 10:1176107.
doi: 10.3389/fmolb.2023.1176107

Variability of the innate immune response is globally constrained by transcriptional bursting

Nissrin Alachkar[1], Dale Norton[1], Zsofia Wolkensdorfer[1], Mark Muldoon[2] and Pawel Paszek[1]*

[1]Division of Immunology, Immunity to Infection and Respiratory Medicine, Lydia Becker Institute of Immunology and Inflammation, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom, [2]Department of Mathematics, University of Manchester, Manchester, United Kingdom

Transcription of almost all mammalian genes occurs in stochastic bursts, however the fundamental control mechanisms that allow appropriate single-cell responses remain unresolved. Here we utilise single cell genomics data and stochastic models of transcription to perform global analysis of the toll-like receptor (TLR)-induced gene expression variability. Based on analysis of more than 2000 TLR-response genes across multiple experimental conditions we demonstrate that the single-cell, gene-by-gene expression variability can be empirically described by a linear function of the population mean. We show that response heterogeneity of individual genes can be characterised by the slope of the mean-variance line, which captures how cells respond to stimulus and provides insight into evolutionary differences between species. We further demonstrate that linear relationships theoretically determine the underlying transcriptional bursting kinetics, revealing different regulatory modes of TLR response heterogeneity. Stochastic modelling of temporal scRNA-seq count distributions demonstrates that increased response variability is associated with larger and more frequent transcriptional bursts, which emerge via increased complexity of transcriptional regulatory networks between genes and different species. Overall, we provide a methodology relying on inference of empirical mean-variance relationships from single cell data and new insights into control of innate immune response variability.

KEYWORDS

transcriptional bursting, burst size, burst frequency, stochastic transcription, telegraph model, innate immunity, toll-like receptor, scRNA-seq inference

Introduction

Transcription of almost all mammalian genes occurs in bursts, during brief and random periods of gene activity. The patterns of temporal mRNA production in a single cell, and the overall mRNA (and protein) distribution in cellular populations, are controlled by transcriptional bursting, namely, via the modulation of *burst size* and *burst frequency* (Raj et al., 2006; Suter et al., 2011; Molina et al., 2013). The innate and adaptive immune responses exhibit extreme variability at the single cell level, in comparison to other tissue systems (Shalek et al., 2013; Shalek et al., 2014; Hagai et al., 2018), where only subsets of cells produce specific effector molecules, and thus are able to respond to pathogen (Avraham et al., 2015; Iakovlev et al., 2021). This apparent level of variability poses a fundamental systems biology question;

how do robust immune responses emerge from this heterogeneous transcriptional bursting process?

Recent advances have demonstrated key insights into regulation of transcriptional bursting. In general, the bursting kinetics are gene-specific and subject to regulatory control via cellular signalling events (Suter et al., 2011; Larson et al., 2013; Megaridis et al., 2018; Wong et al., 2018; Bass et al., 2021) as well as genome architecture and promoter sequences (Dar et al., 2012; Dey et al., 2015; Zoller et al., 2015; Hagai et al., 2018; Ochiai et al., 2020; Einarsson et al., 2022). For example, core promoters control burst sizes, while enhancer elements modulate burst frequency to define cell-type specific (Larsson et al., 2019) or circadian gene expression outputs (Nicolas et al., 2018). Coordinated gene activity has also been shown to regulate mRNA outputs as a function of spatial position during development (Zoller et al., 2018; Hoppe et al., 2020; Wang et al., 2020) as well as temporal immune responses (Robles-Rebollo et al., 2022). The resulting cell-to-cell variability is a consequence of the stochastic processes governing signalling and transcription (Elowitz et al., 2002), but also reflects extrinsic differences between individual cells (Spencer et al., 2009; Adamson et al., 2016; Phillips et al., 2019; Shaffer et al., 2020) or variability of the pathogen in the context of the innate immune response (Avraham et al., 2015). With individual genes exhibiting different levels of stimuli-induced heterogeneity, we are still lacking general understanding of how transcription is regulated at the single cell level.

Toll-like (TLR) receptor signalling constitutes one of the fundamental, evolutionarily conserved innate immune defence mechanisms against foreign threats (Gay et al., 2014; Bryant et al., 2015), yet exhibits substantial cell-to-cell variability (Shalek et al., 2013; Shalek et al., 2014; Lu et al., 2015; Xue et al., 2015; Hagai et al., 2018). We recently demonstrated that this overall TLR response to stimulation (or in general perturbation) is constrained through gene-specific transcriptional bursting kinetics (Bagnall et al., 2020). By utilising single molecule Fluorescent *in situ* Hybridisation (smFISH), we established that the overall mRNA variability is linearly constrained by the mean mRNA response across a range of related stimuli. Variance (and in fact higher moments) of the mRNA distributions have been also shown to be constrained by the mean response in the developing embryo (Zoller et al., 2018). These analyses suggest that complex transcriptional regulation at a single cell level may be globally characterised by mean-variance relationships of gene-specific mRNA outputs, providing new ways to characterise response variability. While quantitative smFISH provides important insights, this approach is often limited by the number of genes, which can be investigated (Raj et al., 2008; Zenklusen et al., 2008; Larson et al., 2013; Lee et al., 2014; Gomez-Schiavon et al., 2017; Bagnall et al., 2018; Bagnall et al., 2020; Bass et al., 2021), therefore further analyses of global gene expression patterns (Larsson et al., 2019; Ochiai et al., 2020) are required to fully understand the underlying regulatory constraints.

Here we utilise scRNA-seq data on innate immune phagocytes stimulated with common TLR ligands, lipopolysaccharides (LPS) of Gram-negative bacteria upstream of TLR4 and viral-like double-stranded RNA (PIC) for TLR3 (Hagai et al., 2018) to investigate the control of single cell gene expression heterogeneity of the innate immune responses. We analyse 2,338 TLR-response genes and demonstrate that they globally follow empirical linear mean-variance relationships, exhibiting a genome-wide spectrum of response variability levels characterised by the slope of the relationship. We show that linear relationships define different modes of individual-gene response modulation with majority of the genes undergoing frequency modulation to TLR stimulation. Mathematical modelling of scRNA-seq count distributions using dynamic stochastic telegraph models of transcription of varied complexity levels, demonstrates that increased response variability is associated with larger and more frequent transcriptional bursts, which emerge via increased regulatory complexity. Finally, we show that linear mean-variance relationships capture evolutionarily differences in response variability across pig, rabbit, rat, and mouse and predict transcriptional bursting modulation between species. Overall, our data demonstrate the utility of empirical mean-variance relationships in providing new insights into control of transcriptional variability in the innate immune response.

Results

TLR-induced mRNA responses exhibit linear mean-variance trends

To globally investigate the control of transcriptional bursting in the TLR system relationships we used existing scRNA-seq data from mouse phagocytes either untreated or stimulated with LPS and PIC for 2, 4 and 6 h (Hagai et al., 2018). The dataset contains unique molecular identifier (UMI) mRNA counts for 53,086 cells and 16,798 genes across 20 experimental conditions including replicates, of which 2,338 genes were identified as TLR-dependent (see Figure 1A for correlation of sample mean and variance across all datasets, and Materials and Methods for data processing). While in general, there is a nonlinear relation between the variance and mean response, in agreement with other analyses (Taniguchi et al., 2010; Dar et al., 2016), the relative variability in the data (captured by a coefficient of variation, i.e., standard deviation normalised by the mean) decreases as the level of response increases (Supplementary Figure S1A). We previously showed that the gene-specific variability can be defined by the slope of the mean-variance relationship (Bagnall et al., 2020). To test this phenomenon globally, for each of the 2,338 TLR-inducible genes, the sample mean (μ) and variance (σ^2) relationship was fitted using robust linear regression ($\sigma^2 = \alpha\mu + \alpha_0$), yielding 2,133 genes with a significant regression slope (p-value < 0.05, Figure 1B). Of those, 1,551 (66% of all TLR-inducible genes) genes, referred here as high confidence genes, were characterised by coefficient of determination $R^2 > 0.6$ (Figure 1C, see also Supplementary Table S1 for list of genes and fitted relationships). Overall, the distribution of fitted slopes across the high confidence genes varied over 3 orders of magnitude, with 1,067 genes (69% of high confidence genes) characterised by slope $\alpha > 1$ and 627 (40%) $\alpha > 3$, indicative of predominant non-Poissonian transcription (where one would expect $\alpha = 1$ and $\alpha_0 = 0$) (Figure 1D). 61 genes (4%) were characterised by $\alpha > 5$ and 28 (2%) by a $\alpha > 10$, highlighting genes with the highest level of expression variability (across a range of TRL responses, Supplementary Figure S1B). Among the high variability genes ($\alpha > 5$) we found C-C motif chemokine ligands (*Ccl 2, 3, 4, 5, 17*; C-X-C motif ligands (*Cxcl 9* and *10*, as well as cytokines including Interleukin 1 α (*IL1a*), *IL1b*, *IL10*, *IL12b* and Tumour Necrosis Factor α (*Tnfa*) (see Figure 1E for individual gene fits). The most

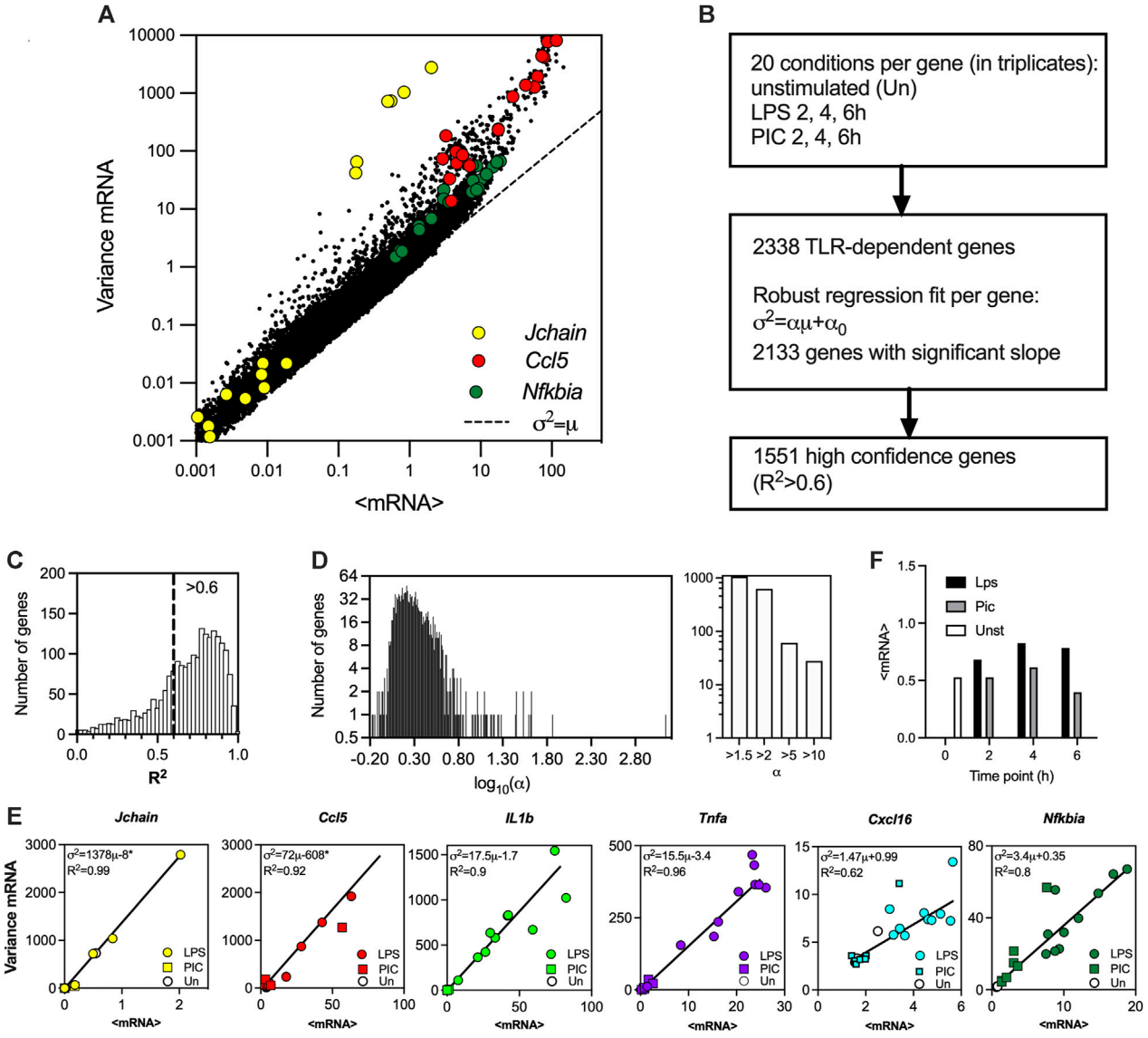


FIGURE 1
TLR-induced transcriptional variability is linearly constrained. **(A)**. Overall variability in the scRNA-seq dataset (Hagai et al., 2018). Shown is the scatter plot of the sample mean (μ) and variance (σ^2) calculated for 2340 TLR-dependent genes across 20 experimental conditions. Data points corresponding to *Jchain*, Ccl5 and *Nfkbia* highlighted in yellow, red, and green, respectively. Broken line indicates $\mu = \sigma^2$ line. **(B)**. Schematic description of the fitting protocol. **(C)**. Histogram of coefficient of determination (R^2) for 2,133 gene fits characterised by a significant regression slope (p-value < 0.05). R^2 = 0.6 broken line corresponds to the high confidence gene cut-off. **(D)**. Distribution of the fitted regression slopes for the 1,551 high confidence gene set. Histogram of the fitted slopes shown on the left. Number of genes with different slope range shown on the right. **(E)**. Fitted mean-variance relationships for a subset of genes. Shown are the individual datapoints (LPS, PIC and unstimulated) as well as fitted regression line with a fitted equation (* denotes statistically significant intercept, p-value < 0.05) and the coefficient of determination (R^2). **(F)**. Mean mRNA counts across treatments (LPS, PIC) and time (0, 2, 4, 6 h) for the 1,551 high confidence genes.

variable gene in the dataset was the immunoglobulin subunit *Jchain* with α = 1,372 (Supplementary Figure S1B), substantially more than the 2nd most variable *Ccl5* (α = 72). While the range of the mRNA output among high confidence genes varies over 3 orders of magnitude (Supplementary Figure S1C), we found that LPS induced more robust activation than PIC in terms of average expression (Figure 1F). The range of the response significantly correlated with the slope of mean-variance relationships across the 1,551 confidence genes (Spearman's rank correlation r = 0.47, Supplementary Figure S1D). This suggests, that at least in part, the slope of the relationship and thus the heterogeneity of individual gene is related to the amplitude of the stimuli-induced response.

Patterns of transcriptional bursting modulation underlie TLR response heterogeneity

Having established the linear relationships relating the gene-specific transcriptional variability to mean expression, we sought to

10.3389/fmolb.2023.1176107

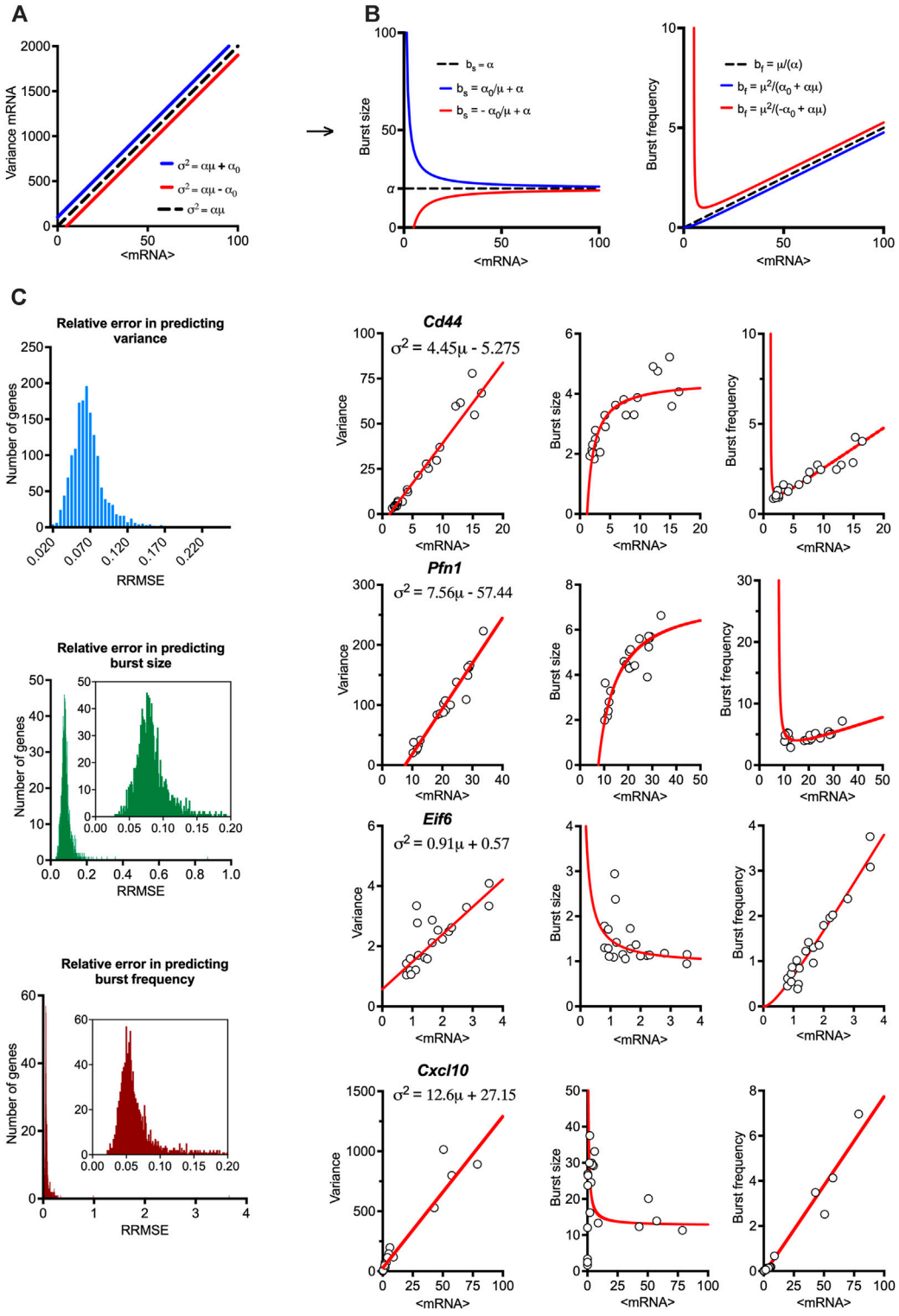


FIGURE 2
Mean-variance relationships constrain transcriptional bursting characteristics. **(A)**. Theoretical burst size and frequency characteristics. (Left) Simulated mean variance relationships with positive (in blue, α = 20, α0 = 100) and negative (in red, α = 20, α0 = −100) intercepts, respectively. (Middle & Right) Derived burst size and frequency modulation schemes for corresponding parameter values calculated using moment estimators. A special case of α = 20, α0 = 0 is shown in broken line. **(B)**. Global modulation of transcriptional busting. Shown is the comparison between fitted mean-variance relationship and derived theoretical burst size and frequency modulation schemes vs. experimental data. Shown is distribution of relative root mean square error (RRMSE) of 1,551 high confidence genes **(C)**. Modulation schemes for *Cd44*, *Pfn1*, *Eif6* and *Cxcl10* genes. Shown is the comparison between theoretical relationships based on fitted mean-variance relationships (in red) and corresponding estimates from data (open circles). Equations for fitted mean-variance relationships highlighted in the top left panel, respectively.

study global properties of transcriptional bursting underlying these trends. We used moment estimators of the underlying scRNA-seq count distributions to calculate bursting characteristics, such that burst size $b_s = \sigma^2/\mu$ (i.e., the Fano factor) and burst frequency $b_f = \mu/b_s$ (Nicolas et al., 2017). These estimators rely on the moments of the mRNA distributions to quantify 'burstiness' as a departure 'nonbursty' Poissonian mRNA production, characterised by $b_s = 1$ and $bf = \infty$ (Raj et al., 2006; Suter et al., 2011; Nicolas et al., 2018; Wong et al., 2018). Given the empirical linear constraint, $\sigma^2 = \alpha\mu + \alpha_0$, the burst size and burst frequency become analytical functions of the mean mRNA expression such that $b_s = \alpha_0/\mu + \alpha$ and $b_f = \mu^2/(\alpha_0+\alpha\mu)$ (Figure 2A). In a special case when $\alpha_0 = 0$, burst size is constant (independent of the mean expression μ) and equal to the slope of the mean-variance line α, while the frequency increases linearly with μ and is proportional to $1/\alpha$ (Bagnall et al., 2020). However, the overall behaviour does depend on the intercept (see Supplementary Figure S1E for sensitivity analyses); for $\alpha_0>0$, the burst frequency converges monotonically to μ/α (i.e., the limiting case for $\alpha_0 = 0$), while the burst size converges to α (from ∞ at $\mu = 0$) as the mean expression μ increases (Figure 2A in blue). For $\alpha_0<0$ (Figure 2A, in red), the relationship can only be defined for $\mu>|\alpha_0|/\alpha$, such that burst size increases monotonically (and converges to α), while the burst frequency has a local minimum for $\mu^* = 2|\alpha_0|/\alpha$ equal to $4|\alpha_0|/\alpha^2$, eventually converging to the limiting case μ/α.

We calculated the theoretical bursting modulation schemes for the 1,551 high confidence genes and compared these to the moment estimators of the burst size and frequency from the data (Figure 2B). We found that the average relative root mean square error (RRMSE, see methods) of the mean-variance fit in relation to data was 0.07% ± 0.02%, where 1,431 genes had an error smaller than 0.1%. In comparison, the average error for the burst size modulation was 0.08% ± 0.03% (with 1,281 genes having an error smaller than 0.1%), while the average error for the burst frequency modulation was 0.07% ± 0.1% (with 1,389 genes having an error smaller than 0.1%). Given their empirical nature, the predicted theoretical trends are in good agreement with the changes of burst size and frequency observed in the data. Profilin 1 (*Pnf1*) and *Cd44* are example genes characterised by intercept $\alpha_0<0$, while the genes encoding eukaryotic translation initiation factor 6 (*Eif6*) and *Cxcl10* had $\alpha_0>0$ (Figure 2C). *Jchain* is an example of a gene with a good mean-variance fit, but one of the poorest fit in terms of bursting frequency, which might be due to limited sample size and its profound variability. Of the 1,551 high confidence genes, 430 genes had a significant intercept (p-value < 0.05) in the regression fit, with 414 characterised by negative and 16 positive intercepts (Supplementary Figure S1F). While intuitively zero intercept is expected (i.e., no expression in untreated conditions), these in part reflect the empirical nature of these trends and the limited sample size, especially for those genes where α_0 is small (in relation to variance), for example, *Cxcl10* (Figure 2C). However, many genes, including *Pnf1* and *Eif6* exhibit substantial basal expression in untreated cells (Bass et al., 2021), resulting in either elevated or reduced variability (in relation to true zero) being captured via non-zero intercept in the regression fit (Bagnall et al., 2020).

Gene-specific bursting exhibits different modes of response modulation

The linear mean-variance relationships reflect the constrained changes of burst size and burst frequency required to regulate response variability as shown in their derived analytical functions of the mean mRNA expression. To understand the modulation of transcriptional bursting, we first calculated fold changes of burst size vs. burst frequency across the range of mean expression calculated for individual response genes (Figure 3A). We found that 1,015 out of the 1,551 high confidence genes exhibit 2 times more fold changes in burst frequency than burst size. This suggests a predominant frequency modulation, in agreement with recent analyses of LPS-induced macrophages (Robles-Rebollo et al., 2022). However, we also found 48 genes exhibiting fold changes in burst size 2 times more than burst frequency, while 389 exhibited comparable modulation of both burst size and burst frequency. To study the transcriptional bursting modulation more systematically, we derived an analytical relationship between the burst size and frequency (independent of the mean mRNA expression) based on the linear constrains (Figure 3B). The general relationship is given by $b_f = \alpha_0/(b_s(b_s-\alpha))$, where α_0 can take positive or negative values. When $\alpha_0>0$, we have an inverse relationship between the burst size and frequency, which asymptotically approaches zero, as the burst size approaches infinity. It is also worth mentioning that, in this case, the function is undefined for values of burst size smaller than or equal to α (Figure 3B, in blue), reflecting a biological limit of burst size and frequency for genes following this modulation trend. We found that 315 genes (out of the 1,551 high confidence genes) exhibited such an inverse relationship, with all genes exhibiting higher frequency than burst size modulation (see Figure 3C for specific genes and Figure 3D and Supplementary Table S2 for global analysis). For the case when $\alpha_0<0$, linear constrains define a non-monotonic relationship between the burst size and frequency on the interval (0,α) with a local minimum at $b_s^* = \alpha/2$, and frequency diverging to infinity as burst size tends towards α or is close to 0 (Figure 3B, in red). From the case $\alpha_0<0$, three patterns of bursting modulation can be distinguished; the burst frequency and size exhibit either inverse relationship, where the frequency increases and burst size decreases (for $b_s < b_s^*$) or concurrent increases ($b_s > b_s^*$). In addition, we define a U-shape relationship where the inverse or concurrent relationship is possible (i.e., $b_{s\,max} > b_s^*$ and $b_{s\,min} < b_s^*$, per gene), but changes occur only close to the minimum of the function (such that $b_s \approx b_s^*$), unlike other relationships. This mode allows greater burst size modulation (in 218 of 767 genes) comparing to other modes (19 genes, Figures 3C, D). We found that out of the 1,236 genes characterised by $\alpha_0<0$, most genes (999) exhibited predominant frequency modulation following either a U-shape or a concurrent relationship. It is worth mentioning that all 7 genes confirming an inverse trend showed predominant burst size modulation. Overall, these analyses demonstrate different modes of the transcriptional bursting modulation of TLR-stimulated genes, albeit with predominant regulation via burst frequency.

10.3389/fmolb.2023.1176107

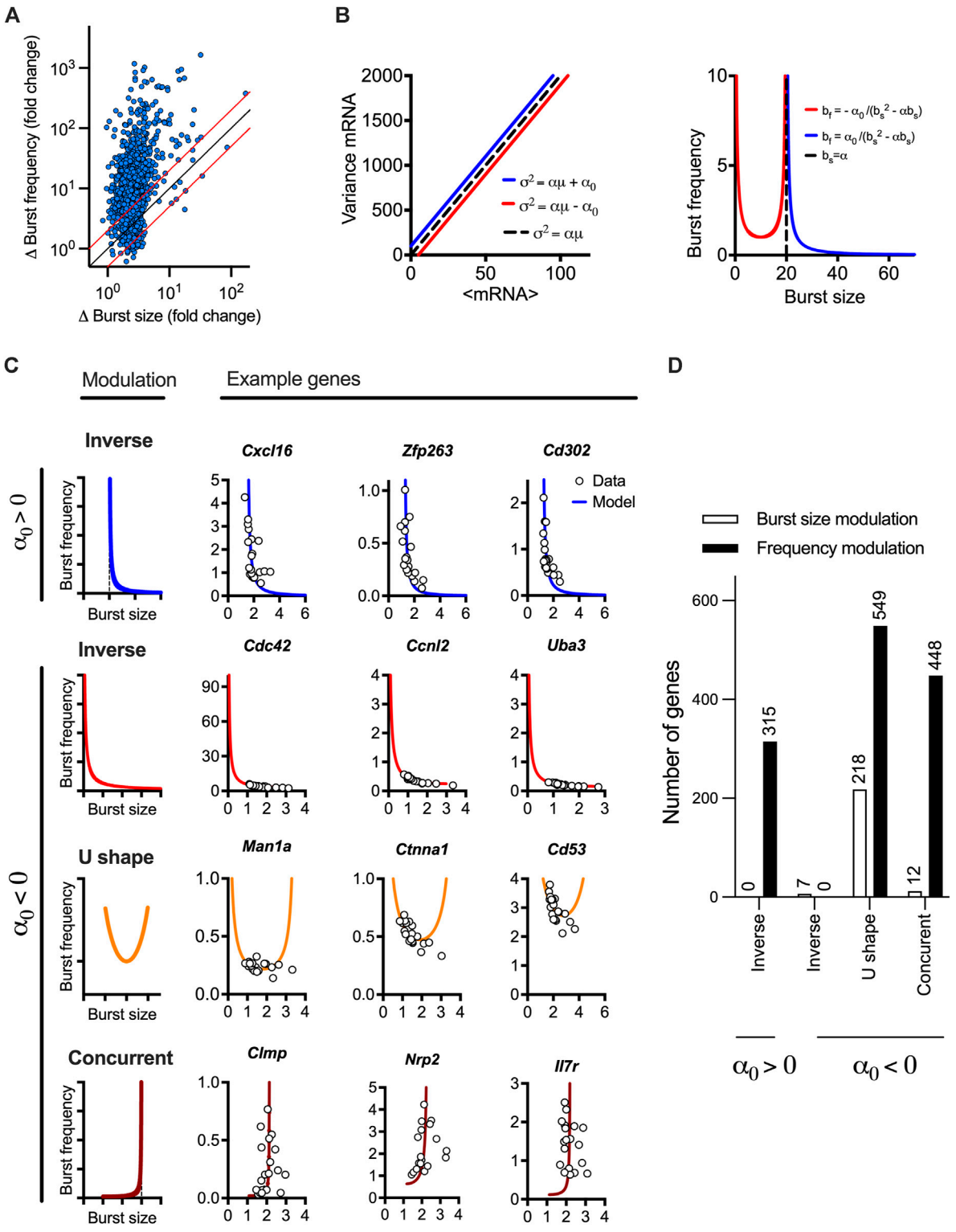


FIGURE 3
LPS-induced gene expression undergoes different modes of transcriptional bursting. **(A)**. Relative changes of burst size and burst frequency. Shown is the relative fold change of burst size and frequency calculated across the individual range of mean expression for 1,551 high confidence genes (in blue circles). Identity line depicted in black, two-fold change highlighted in red. **(B)**. Theoretical relationship between burst size and frequency. (Left) Simulated mean variance relationships with positive (in blue, $\alpha = 20$, $\alpha_0 = 100$) and negative (in red, $\alpha = 20$, $\alpha_0 = -100$) intercepts, respectively. (Right) Burst size and frequency modulation schemes for corresponding parameter values calculated using moment estimators. A special case of $\alpha = 20$, $\alpha_0 = 0$ shown in broken line. **(C)**. Modulation of burst size and frequency across a range of individual genes. Shown are inverse relationship ($\alpha_0 > 0$) in blue as well as inverse, U-shape and concurrent relationships ($\alpha_0 < 0$). Relationship predicted from linear constraints in solid lines and corresponding estimates from experimental data in open circles. U-shape numerically defined as maximum burst size value > $\alpha/2$ and minimum burst size value < $\alpha/2$ across conditions. **(D)**. Prevalence of different modulation schemes across 1,551 high confidence genes. Definition of the mode as in C, dominant modulation defined by absolute difference in the burst size vs. frequency changes across the respective range of mean expression (as in A).

10.3389/fmolb.2023.1176107

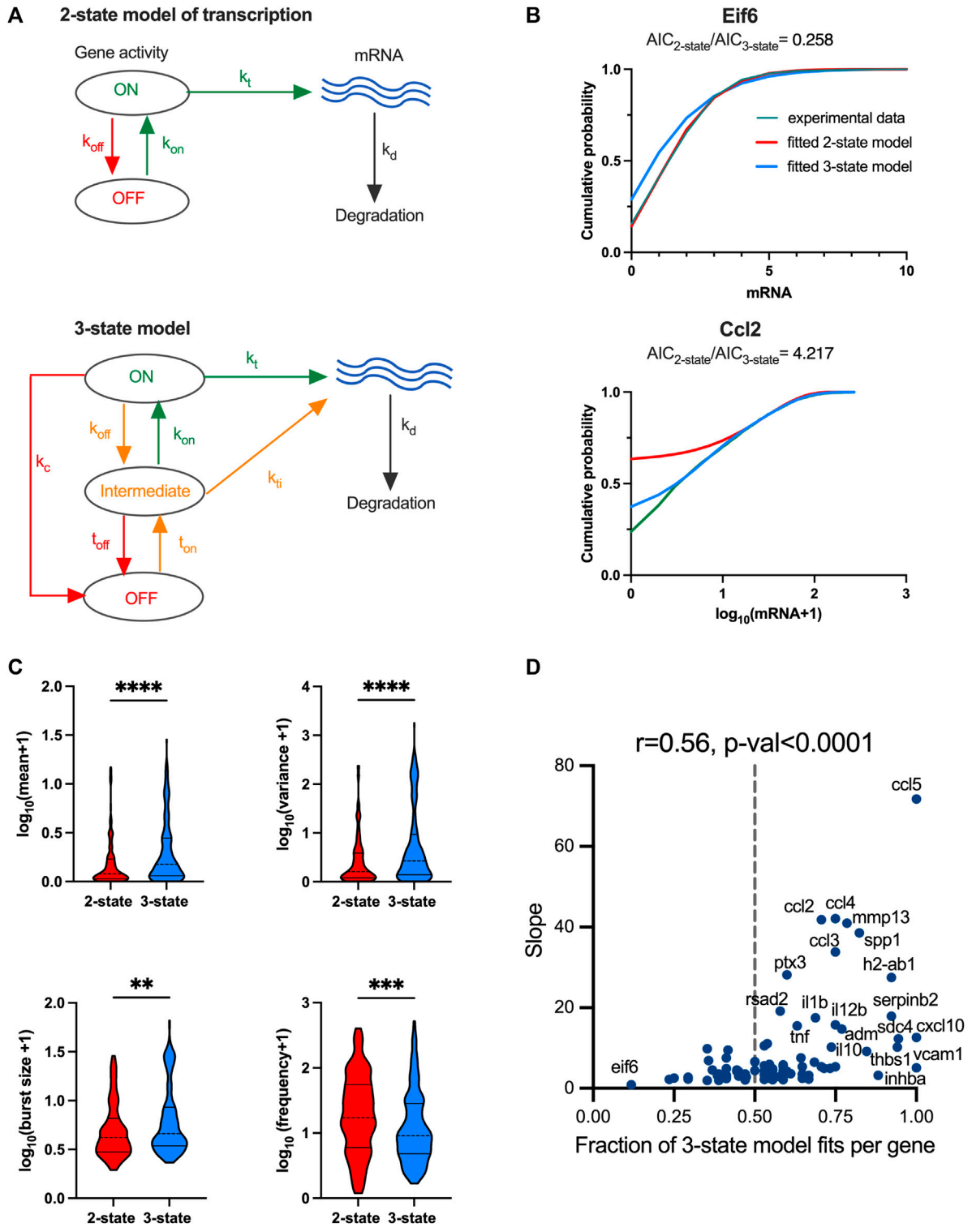


FIGURE 4
TLR response variability is associated with regulatory complexity. **(A)**. Schematic representation of the 2-state and 3-state models of transcription. **(B)**. Comparison between the fitted and measured mRNA counts distributions. Shown are cumulative probability distribution of data (in green) vs. the corresponding 2-state and 3-state stochastic model fits (in red and blue, respectively) for representative condition for *Eif6* (PIC, 4h, replicate 3) and *Ccl2* (LPS, 2h, replicate 2) genes. **(C)**. Analysis of transcriptional bursting across high coverage genes and conditions fitted by 2-state vs. 3-state models. Shown is the comparison between best fit 2-state and 3-state models in terms of mean mRNA expression, variance, burst size and frequency from experimental data. Best fit defined by $AIC_{best\ model}<0.5AIC_{2nd\ best}$ (from Supplementary Figure S3B). Burst size and frequency calculated per condition using moment estimators. Statistical significance assessed with Mann-Whitney test (** p-value < 0.01, **** p-value < 0.0001). **(D)**. Relationship between slope of the mean-variance relationship and fraction of 3-state model fits for high coverage genes. Fraction of 3-state model fits per gene defined by the number of conditions with $AIC_{3\text{-}state\ model}<AIC_{2\text{-}state}$ over all conditions per gene. Broken line indicates 0.5, r denotes Spearman correlation coefficient.

Increased response variability is associated with complex transcriptional regulation

The distribution of fitted regression slopes varying over 3 orders of magnitude demonstrate a wide range of response variability among individual TLR-induced genes (Figure 1D). While we have demonstrated that individual genes exhibit different modes of transcriptional bursting characteristics to regulate responses to stimulation, we wanted to understand the control of variability in the system more mechanistically. A well-established mathematical description of mRNA production involves a 2-state telegraph model (Figure 4A), where gene activity changes randomly between "off" and "on" states, with mRNA transcription occurring in the "on" state (Raj et al., 2006; Zenklusen et al., 2008; Suter et al., 2011; Nicolas et al., 2018). The associated parameters are gene activity rates (k_{on} and k_{off}) as well as rate of mRNA transcription (k_t) and degradation (k_d) (Nicolas et al., 2018). Although the 2-state telegraph model has been widely used in the past to model mRNA count data, more complex structures are often required to capture additional complexity associated with multiple regulatory steps, combinatorial promoter cycling and transcriptional initiation (Harper et al., 2011; Zoller et al., 2015). We previously showed that heterogenous *Il1β* mRNA transcription requires more regulatory steps than that of *Tnfα* (Bagnall et al., 2020). We therefore hypothesised that TLR response variability is linked with the complexity of the transcriptional regulation. To test this hypothesis, we introduced a 3-state stochastic model, which assumes sequential promoter activation between "off", "intermediate" and "on" states, equivalent to promoter cycling (Harper et al., 2011; Zoller et al., 2015), with transcription occurring in the "intermediate" (I) state as well as in the "on" state, characterised by 5 transition rates (t_{on}, t_{off}, k_{on}, k_{off} and k_c), 2 transcription rates (k_{ti} and k_t), and a degradation rate k_d (Figure 4A).

We first used a profile likelihood approach (Vu et al., 2016; Larsson et al., 2019) to fit the measured scRNA-seq count distributions assuming steady state kinetics of the 2-state model (the so called Beta-Poisson model) for the 1,551 high confidence genes, each across 20 treatment datasets (Supplementary Table S3). Values of kinetic parameters were inferred for 7,804 of 31,020 datasets (~25% across 1,519 genes), which in general corresponded to genes characterised by larger expression, in comparison to those that failed to fit (Supplementary Figure S2A). The fitted parameter values (k_{on}, k_{off} and k_t, expressed in units per degradation half-life) varied over 3 orders of magnitude across all genes and datasets (Supplementary Figure S2B). In general, gene inactivation rates (k_{off}) were greater than activation rates (k_{on}) (Supplementary Figure S2C), consistent with intermittent transcriptional kinetics (Suter et al., 2011; Dar et al., 2012; Larsson et al., 2019). While the Beta-Poisson model explicitly assumes a steady-state (and does not make any assumptions about mRNA half-life), we wanted to account for the underlying dynamical stochastic processes and corresponding temporal mRNA production and decay (Gomez-Schiavon et al., 2017). However, it was not computationally feasible to fit all genes across all scRNA-seq datasets, we therefore identified on a subset of 99 high confidence genes for which at least 10 datasets were fitted using a Beta-Poisson model (Supplementary Figure S2D). Of these, 96 had an existing measurement of mRNA half-life (which is required for dynamical model fitting) in LPS-stimulated bone marrow derived macrophages (Hao and Baltimore, 2009; Kratochvill et al., 2011) or other cell models (Maurer et al., 1999; Raghavan et al., 2002; Park et al., 2004; Sharova et al., 2009; Kambara et al., 2014; Payne et al., 2014; Martin et al., 2017; Zainol et al., 2019) (see Supplementary Table S3 for specific values). The resulting 96 high coverage genes included 51 of 100 most variable genes (as defined by the fitted regression slope) and 60 of 100 most expressed genes including chemokine family *Ccl5*, *Ccl4*, *Ccl3*, *Ccl2* as well as *IL1b* and *TNFa* (Supplementary Figure S2D–F, see Supplementary Table S3 for a list of genes and fitted relationships).

We used a genetic algorithm to fit dynamical 2-state and 3-state stochastic models across 20 individual datasets (LPS and PIC stimulation at 0, 2, 4, 6 h time-course across replicates) for the 96 high coverage genes (see Material and Methods). We then applied the Akaike information criterion (AIC) (Akaike, 1973) incorporating the penalty for model complexity, to select the simplest (i.e., best fit) model that accurately fitted the measured mRNA distributions per condition, noting that the lower AIC value corresponds to the better model fit. In general, we found that Beta-Poisson model, the least constrained model, fitted better than dynamical models (805 out of 1,210 conditions (i.e., treatment and replicates) favoured Beta-Poisson model based on their AIC values, Supplementary Figure S3A, B). The more constrained dynamical 2-state model provided a best fit for 170 conditions, while the 3-state model best captured 235 conditions (and 30 and 57, respectively when using a more stringent criterion of two-fold AIC change, Supplementary Figure S3B). When comparing 2-state with 3-state model directly and assuming a two-fold AIC change between the two models, there were 141 out of 1,507 conditions that favoured the 2-state model, while the opposite was true for 266 conditions (see Supplementary Figure S3C for other thresholds). For example, 2-state model recapitulated PIC-treated *Eif6* mRNA count distribution (at 4 h) better than a 3-state model, as reflected by the $AIC_{2\text{-}state} < AIC_{3\text{-}state}$, this to some extent reflects the fact that although generally more accurate, the 3-state model is also more difficult to fit by the genetic algorithm. In turn, the 3-state model better recapitulated the LPS-treated *Ccl2* distribution (at 2 h) spanning almost over 3 orders of magnitudes (Figure 4B). The number of 2-state-and 3-state model fits was not strongly related to the treatment, time point or in fact biological replicates, although LPS had 155 conditions more fitted with 3-state than 2-state model (Supplementary Figure S3D).

The 141 2-state model fits were characterised by $k_{on} = 0.02 \pm 0.01$ min^{-1} (half-time of 35 min) on average, and off rates averaging $k_{off} = 0.74 \pm 0.25$ min^{-1} (half-time of 1 min), with average transcription rate $k_t = 1.23 \pm 4.44$ mRNA min^{-1}, indicative of 'bursty' kinetics (Supplementary Figure S4A). The '*on*' rate showed significant positive correlation with the variance of the corresponding count distributions ($r = 0.48$), demonstrating that a faster '*on*' switch contributes towards increased response variability. The 266 3-state model fits were also characterised by relatively slow average '*on*' rates ($t_{on} = 0.036 \pm 0.13$ min^{-1} and $k_{on} = 0.33 \pm 0.32$ min^{-1}) in relation to the '*off*' rates ($t_{off} = 0.74 \pm 0.26$ min^{-1}, $k_{off} = 0.44 \pm 0.36$ min^{-1} and $k_c = 0.50 \pm 0.36$ min^{-1}, Supplementary Figure S4B). The mRNA count variance was correlated positively with t_{on} rate (i.e., transition to intermediate state, $r = 0.33$) as well as with transcription rates in 'on'

10.3389/fmolb.2023.1176107

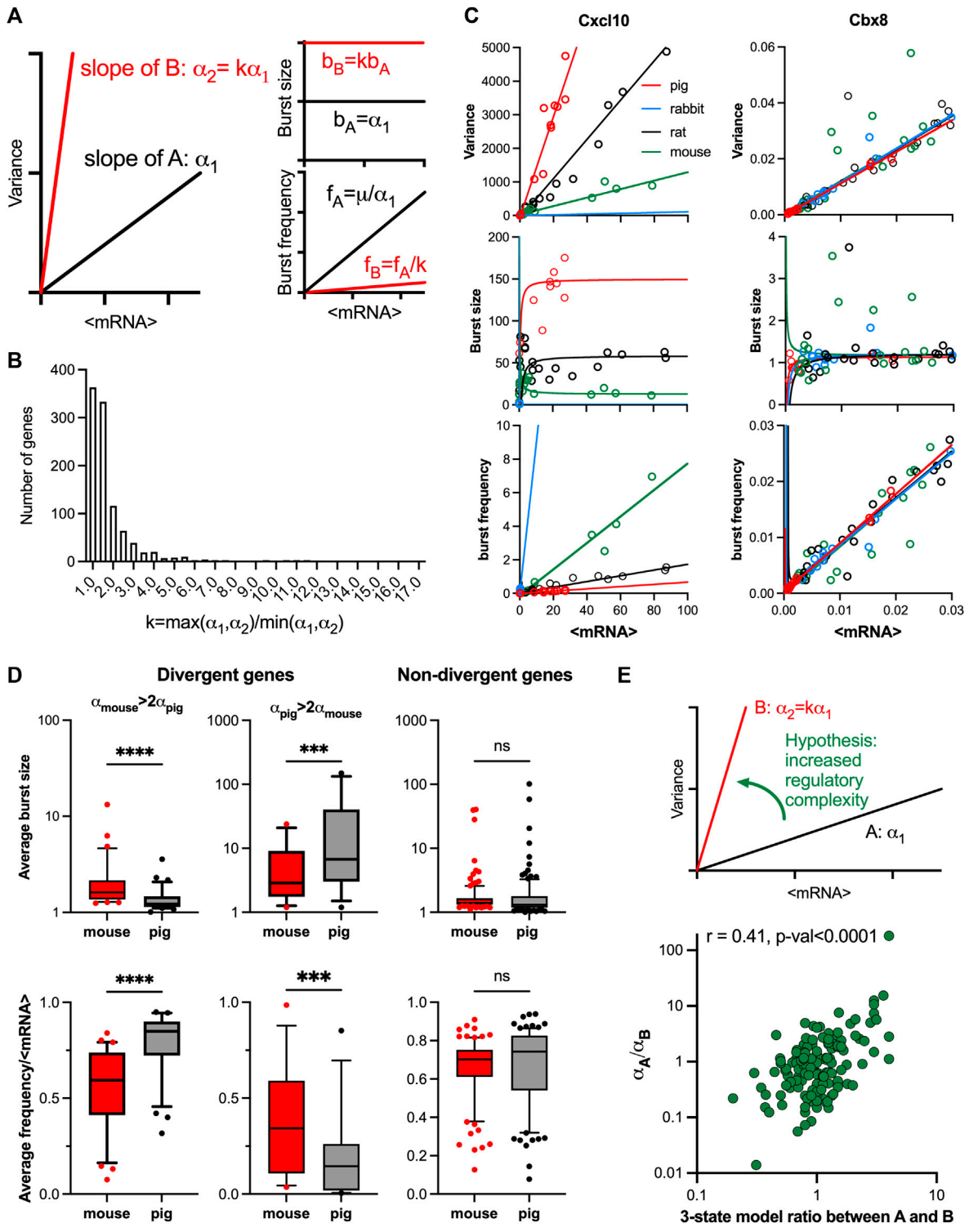


FIGURE 5
Evolutionary control of TLR response variability. **(A)**. Schematic representation of response variability during evolution for putative species A and **(B)**. Shown are mean variance relationships corresponding to slopes (α1 and α2 = kα1) and the predicted burst size **(B)** and frequency **(F)** modulation schemes for corresponding parameter values calculated using moment estimators. **(B)**. Histogram of the slope ratio k calculated for the 169 orthologue genes across all pairwise comparisons between mouse, rat, rabbit and pig. k = max (α1,α2)/min (α1,α2), where α1 and α2 denote slopes of the fitted mean-variance relationships for each pair of species per gene. **(C)**. Modulation schemes for Cxcl10 and Cbx8 genes. Shown is the comparison between theoretical relationships based on the fitted mean-variance relationships (in solid lines, colour-coded by species) and corresponding moment estimates for burst size and frequency from experimental data (circles). **(D)**. Analysis of burst size and frequency for divergent and non-divergent mouse and pig TLR-response genes. Shown are box plots of average burst size and mean-normalized frequency per gene stratified into divergent (αmouse> 2αpig or αpig>2αpig) and complementary non-divergent groups (31, 15 and 123 orthologue genes, respectively). Statistical significance assessed with a paired Wilcoxon test (**** *p*-value < 0.0001, *** *p*-value < 0.001, ns not significant). **(E)**. Change of variability between species is associated with regulatory

(*Continued*)

FIGURE 5 (Continued)
complexity. Top: Schematic representation of the hypothesis. Bottom: Relationship between the slope ratio (αA/αB) estimated for 146 pairwise comparisons between 28 fitted orthologue genes for mouse, rat, rabbit and pig; and the corresponding ratio between species A and B of the number of conditions per gene with 3-state model fitting better than 2-state model. Absolute difference in AIC of the two models was used for model selection. Shown is the Spearman correlation coefficient r and a p-value for $r > 0$.

and 'intermediate' states (r > 0.4). In comparison to the 2-state model, the transcription rate in the 'on' state was significantly higher ($k_t = 7.63 \pm 13.05$ mRNA min^{-1}) indicative of larger burst sizes (Supplementary Figure S4C, D).

We then asked if the level of variability is linked with the model complexity. We found that scRNA-seq count distributions fitted with the 3-state model were characterised by greater variability than those corresponding to the 2-state model (see Figure 4C; Supplementary Figure S4D for less stringent model selection thresholds). In agreement, the 3-state-model fits were associated with significantly larger burst size and lower burst frequency than that of the 2-state model fits, consistent with more heterogenous bursting kinetics across the relevant conditions. Finally, we analysed model selection across individual high coverage genes rather than corresponding conditions; we found the fraction of conditions explained by one model changes between individual genes (e.g., 3-state model fitted 3 out of 20 for *Eif6*, 10 out of 20 for *Ccl5* and all conditions for *Vcam1* Figure 4D). Our interpretation of this is that as the mRNA responses increase, a more complex regulatory structure is required to capture the underlying distribution. We found that, for the high coverage genes, the fraction of conditions explained by the 3-state model correlated well ($r = 0.56$, p-value < 0.0001) with the slope of mean-variance relationship, and thus response heterogeneity (Figure 4D). Overall, this demonstrates that while increased heterogeneity involves larger and infrequent bursts (in comparison to homogenous responses), this is underlined by increased complexity of the transcriptional regulatory network.

Linear relationships capture evolutionary changes of response variability

Previous work highlighted the relationship between evolutionary response divergence of innate immune genes and their cell-to-cell variability, with highly divergent genes exhibiting more variability (Hagai et al., 2018). However, the changes in patterns of transcriptional bursting during evolution is still poorly understood. We proposed that by comparing the linear mean-variance relationships across species, the variations in transcriptional bursting patterns that develop through evolution could be better understood. Specifically, if the evolutionary changes in response variability can be captured by a fold-change k in the slope of the relationship, then the increased variability is predicted to be due to increased burst size and reduced burst frequency by a factor k, respectively (Figure 5A).

The relationship between the mean and variance of the single cell mRNA counts was studied in data for four mammalian species from Hagai *et al.* (2018): mouse, rat, pig, and rabbit, in cells either untreated or treated with LPS or PIC for 2, 4 and 6 h (see methods and Supplementary Table S4 for species specific number of conditions per gene ranging from 12 to 21). We found that from the 2,338 LPS response genes, a subset of 218 genes with one-to-one orthologues showed response to treatment in all four species (Supplementary Figure S5A). 78% of fitted mean-variance relationships for the 218 genes were characterised by $R^2 > 0.6$, including 102 genes in all four species and 169 in at least three species. To characterise the divergence in response variability we performed species pairwise comparison between the fitted regression slopes of the 169 genes subset (Supplementary Table S5). Out of this subset 21 genes including chemokines *Ccl2*, *Ccl4*, *Ccl5* and *Cxcl10* (Figure 5B; Supplementary Figure S5B), had all 6 possible pairwise comparisons showing significant differences, indicating divergence in TLR response variability between each of the two species. 5 significant FDR values (difference in three out of four species) were obtained for 49 genes including chemokines *Ccl20*, *Ccl3*, *MMP9* (Supplementary Figure S5B) and cytokines *Il1a*, *Il10* and *Il27* indicating significant differences in response variability. On the other hand, no significant differences were obtained between any of the four slopes in 7 genes, including a transcriptional repressor Chromobox Protein Homologue 8 (*Cbx8*, Figure 5B). In agreement, a distribution of slope ratios calculated across all pairs of species for the 169 genes (Figure 5C and Supplementary Table S6) revealed 49 pairs with $k > 5$ and 258 pairs with $k > 2$, indicating substantial changes of the response variability between species, including the chemokine and cytokine genes. Conversely, 54% of slope ratios (549 out of total 1,014 genes) were smaller than 1.5, indicative of conserved variability. The inflammatory chemokines were shown previously to rapidly evolve in mammals and other vertebrates with clear differences in expression between closely related species (Nielsen et al., 2005; Haygood et al., 2010). Moreover, gene duplication of the CC chemokine ligands can result in different copy numbers of these genes between individuals (Nomiyama et al., 2010), further increasing the divergence in expression. Importantly, our analyses specifically capture changes of response variability and suggest a statistical relationship of these changes with the generic evolutionary divergence (see Materials and Methods) of gene expression response (Supplementary Figure S5C).

To validate the predicted changes in transcriptional bursting during evolution (Figure 5A), we first calculated the theoretical modulation schemes for all the 169 evolutionary genes across species and compared these to the moment estimators of the burst size and frequency from the data (Supplementary Figure S5D). We found that the average RRMSE of the mean-variance fit in relation to data was 0.06% ± 0.05% across all species, where 90% genes had an error smaller than 0.1%. In comparison, the average error for the burst size predictions was 0.08% ± 0.05%, while the average error for the burst frequency predictions was 0.05% ± 0.04%. The predicted theoretical trends are in good agreement with the observed changes of burst size and frequency. For example, *Cxcl10* exhibits concurrent changes of the burst size and frequency, the level of which is determined by the slope of the relationships, while *Cbx8* exhibits the same modulation

across species (Figure 5C). In addition, our predictions of species-specific modulation scheme are based not only on the slope α, but also the mean-variance intercept, which we previously showed may affect the bursting relationships (Figure 2A and Supplementary Figure S1E). We therefore investigated if the difference of the slopes alone is sufficient to predict modulation of bursting characteristics across species (Figure 5A). We stratified the 169 orthologous genes into divergent and non-divergent subsets, with the divergence threshold defined by a 2-fold change in the slope of the mean-variance relationships. The divergent subset included 31 genes exhibiting higher slope in mouse, and 15 in pig (Supplementary Figure S5E). We found that divergent genes, associated with increased response variability, exhibited significantly higher average burst sizes (as calculated across all corresponding conditions) and reciprocally lower normalised burst frequency when compared between the two species (Figure 5D). In contrast, the non-divergent genes showed no significant differences in the burst size or normalized frequency, as predicted by the linear constraints. Interestingly, we also observed significant differences in the average expression between the divergent genes group, opposing to the non-divergent group (Supplementary Figure S5F).

We then asked if the increased variability in gene expression between species was associated with changes of regulatory complexity (Figure 5E). Following previous methodology, we selected 28 orthologue genes from the subset of 96 of high coverage genes in mouse and used a genetic algorithm to recapitulate scRNA-seq count distributions with dynamical 2-state and 3-state models (see Materials and Methods and Supplementary Table S6 for details of the analysis). We then calculated the fold change in the number of conditions (per gene) fitted with 3-state models across all pairwise comparisons of the four species. We found that this fold change correlated (Spearman's r = 0.41, $p < 0.0001$) with the ratio of the slopes between the corresponding linear relationships, such that the transition to a higher slope was associated with increased number of 3-state model fits across corresponding conditions (Figure 5E). Overall, this demonstrates that evolutionary increases in TLR response variability are associated with increased regulatory complexity, resulting in larger and less frequent transcriptional bursting kinetics.

Discussion

Transcription is inherently a stochastic process leading to heterogeneity in cell-to-cell mRNA levels, which has been studied from the inception of systems biology (Paulsson, 2004). The most recent advances suggest the existence of fundamental constraints governing the heterogeneity of gene expression, which rely on the scaling between the variance and mean of the mRNA response distribution (Dar et al., 2016; Zoller et al., 2018). In particular, we previously developed an approach relaying smFISH a comparative analyses of noise across many immune-related conditions (i.e., treatments, doses and times, etc.), which showed that the overall mRNA variability is linearly constrained by the mean mRNA response across a range of immune-response stimuli (Bagnall et al., 2020). However, these approaches were typically limited by the number of genes considered, not allowing to generalise the observations to the genome-wide scale. Here, utilising an existing scRNA-seq data on the evolutionary-conserved innate immune signalling (Hagai et al., 2018), we perform global analysis of the TLR gene expression response variability and underlying transcriptional bursting. We demonstrate that cell-to-cell variability can be empirically described by a linear function of the population mean across a genome. Based on this, we develop a methodology, relying on statistical modelling of linear mean-variance relationships from single-cell data, that provides a simple yet meaningful way to understand regulation of cellular heterogeneity. We demonstrate that (1) The response heterogeneity of a gene can be defined as the slope of the mean-variance line across >1,500 individual response genes. High variability genes include chemokines and cytokines such as CCL family, while other functional genes are more homogenous, in agreement with previous work (Hagai et al., 2018). (2) The changes in heterogeneity between species can be described by the change in the slope of the corresponding mean-variance lines, providing insights into the evolutionary control of TLR response variability. (3) The linear relationships determine the underlying transcriptional bursting kinetics, revealing different regulatory modes in response to stimulation and through evolution. (4) Application of dynamical stochastic models of transcription demonstrates a link between the variability and the regulatory complexity, with complexity facilitating heterogeneity via larger and less frequent transcriptional bursting kinetics.

While, in general the available sequencing data are subject to measurement noise (Luecken and Theis, 2019), and often restricted by the number of data points available, the mean-variance relationships fitted 1,551 genes - 66% of total 2,338 TLR-inducible genes in primary murine phagocytes across 20 experimental datasets (Figure 1). In comparison, out of the 218 genes with one-to-one orthologues between mouse, rat, rabbit and pig, 78% fitted mean-variance relationships despite the number of datapoints being limited to 12 (Figure 5). Fit quality was reflected in the low mean squared errors between the fitted trends and data, providing good support for the observed phenomenon. We found that 430 relationships (out of 1,551 murine fits) were characterised by statistically significant intercept (α_0). While intuitively zero intercept is expected (i.e., no expression in untreated conditions), for some genes, this may reflect the empirical nature of these trends, especially for those with small intercept (in relation to variance), for example, *Cxcl10* (Figure 2C). However, we found that many genes with non-zero intercept fits were associated with substantial basal expression in untreated cells, which was also observed previously for the more quantitative smFISH data (Bagnall et al., 2020). Basal expression of the related gene targets has been shown to exhibit different bursting kinetics and mode of regulation from the inducible expression (Bass et al., 2021), which in part may explain the fitted non-zero intercepts for a subset of genes. For $\alpha_0 = 0$, linear constraints essentially imply that the burst size must be constant (and equal to the slope of the mean-variance line), while the frequency undergoes modulation with the population mean changes in response to stimulation. This is in general agreement with recent analyses demonstrating a role of frequency in regulation of LPS-induced macrophages (Robles-Rebollo et al., 2022) or stimulation (Larson et al., 2013; Fukaya

et al., 2016; Nicolas et al., 2017; Hoppe et al., 2020; Luo et al., 2022). However, the presence of non-zero intercepts in linear fits extends the variety of modulation schemes, including a subset of genes exhibiting burst size modulation (Figure 3). For instance, a positive intercept is associated with an inverse relationship between the burst size and frequency, while a negative intercept may imply concurrent burst size and frequency changes. As with the mean-variance relationships, the predicted modulation schemes are generally in good agreement with the data in terms of the mean-squared error. Notably, we demonstrate that our methodology can be extended to capture evolutionary differences between species. While gene expression divergence between species has been previously measured in terms of the population response (Nourmohammad et al., 2017), the slope of the linear relationships captures the specific differences in TLR response variability through evolution (Figure 5). We demonstrate that the evolutionary change of the variability can be described as a ratio k between the slopes of the corresponding mean-variance fits, which theoretically implies reciprocal scaling of the burst size and frequency also by k. Analysis of the 218 TLR orthologue genes indeed demonstrates that responses of divergent genes are controlled by reciprocal changes of burst size and frequency, while non-divergent genes show the same characteristics across species. Interestingly, we found that within each pair of species, divergent genes exhibited different changes of variability suggesting complex evolutionary traits (e.g., 31 genes exhibiting higher variability in mouse than in pig, and 15 in pig vs. mouse). Our current analyses also suggest that the slope of the mean-variance relationship, at least in part is related to the amplitude of the gene response (Supplementary Figure S1D), suggesting that more inducible genes exhibited increased variability. It would be important to better understand how variability of particular response genes evolved between different species, in the context of their sequence dissimilarities (Nielsen et al., 2005; Haygood et al., 2010; Nomiyama et al., 2010; Einarsson et al., 2022) as well as epigenetic (Lind and Spagopoulou, 2018) and signalling components (Brennan and Gilmore, 2018) of the TLR signalling between species.

We used stochastic models of transcription to better understand regulation of transcriptional bursting (Figure 4). A typical representation involves a 2-state telegraph model, where gene activity changes randomly between “off” and “on” states, facilitating mRNA transcription (Raj et al., 2006; Zenklusen et al., 2008; Suter et al., 2011; Nicolas et al., 2018). However, more complex structures are often used to capture complexity associated with multiple regulatory steps, combinatorial promoter cycling and transcriptional initiation (Harper et al., 2011; Rybakova et al., 2015; Zoller et al., 2015; Yang et al., 2022). We hypothesised that TLR response variability is linked with the complexity of the transcriptional regulation. We introduced a 3-state stochastic model, which assumed a sequential activation between “off”, “intermediate” and “on” states, equivalent to promoter cycling (Harper et al., 2011; Zoller et al., 2015). First, we used a computationally efficient Beta-Poisson model, a steady-state approximation of the 2-state telegraph model, which has previously been used to fit scRNA-seq distributions (Larsson et al., 2019; Luo et al., 2022). However, this model does not take into account the dynamical nature of the process (measurements at 0, 2, 4 and 6 h) and the mRNA half-life with many genes peaking early after stimulation (Hao and Baltimore, 2009). We therefore used a genetic algorithm to fit the theoretical temporal count distributions at specific times to the measured scRNA-seq data using the dynamical 2-state and 3-state models. Based on the Beta-Poisson fits, we selected 96 high coverage murine response genes (and 28 orthologue genes for species analyses), which have existing estimates of mRNA half-life in LPS-stimulated bone marrow derived macrophages (Hao and Baltimore, 2009; Kratochvill et al., 2011) or other cell models. Our current understanding of TLR signalling suggest that due to endotoxin resistance and desensitisation (Buckley et al., 2006; Morris et al., 2014; Kalliara et al., 2022), the regulatory network, and thus model structures and parameters, are time-varying (Wang et al., 2018). For example, previous work show that stability of TLR target genes are regulated in response to stimulation, and also may vary between treatments (Hao and Baltimore, 2009). However, due to limited availability of the data as well as substantially increased computational complexity when considering non-stationary processes (Shand and Li, 2017), we did not incorporate those effects in our models. In general, the measurement of relevant half-lives over times-scales of different stimulation protocols would allow more accurate fits and ultimately better understanding of the influence of time-varying parameters in the system in the future. In our current fitting protocols we treated each data time-point (and replicate) separately, which also allowed more efficient algorithm implementation to fit 1,507 mouse, and 1,079 orthologue conditions. We then used the AIC method (Akaike, 1973) to compare the models considered, including a penalty for complexity, and select the one that fitted the measured mRNA distributions most accurately. The results demonstrated that a large subset of genes and conditions fitted a dynamical 3-state model better than the 2-state model. We found that the fraction of conditions explained by the 3-state model correlated well ($r = 0.56$, p-value < 0.0001) with slope of the mean-variance relationship, and thus response heterogeneity, for the high coverage murine genes (Figure 4). Similarly, the increased complexity was associated with evolutionary changes of response variability between species (Figure 5). In general, we found that increased regulatory complexity facilitated larger response variability through increased burst sizes and reduced frequency of transcriptional bursting (Figure 4D), while scRNA-seq count variance exhibited correlations with transcription rates and ‘on’ rates. A better understanding of the relationships, and in particular mechanistic basics for controlling gene-specific slopes (i.e., response variability) as well as their sensitivity to pharmacological perturbation and infection and disease state, will require further detailed investigations (Robles-Rebollo et al., 2022). Nevertheless, we believe that our methodology, relying on the inference of mean-variance relationships, provides new insight into regulation of single-cell variability of innate immune signalling and will be applicable to other gene expression systems, including prominent stochastic regulation of adaptive immunity (de la Higuera et al., 2019; Iakovlev et al., 2021).

Materials and Methods

Analysis environment

Computational analysis was performed using Python v3.8.2 in a 64-bit Ubuntu environment running under Windows Subsystem for Linux (WSL) 2 and using the conda v4.8.3 package manager. Relevant packages were NumPy v1.19.1 (Van Der Walt et al., 2011), pandas v1.0.5 (Reback et al., 2020), Scanpy v1.5.1 (Wolf et al., 2018), scikit-learn v0.23.1 (Pedregosa et al., 2011), SciPy v1.4.1 (Virtanen et al., 2020) and statsmodels v0.11.1 (Seabold and Perktold, 2010) for processing and Matplotlib v3.2.1 (Hunter, 2007) and seaborn v0.10.1 (Waskom et al., 2020) for visualisation. Robust linear regression models and Benjamini-Hochberg false discovery rate (FDR) correction was performed in statsmodels. Coefficient of determination (R^2) scores were calculated using the metrics module of scikit-learn.

Acquisition and processing of mRNA count data

mRNA count data associated with the study by Hagai *et al.* (2018) were downloaded from the Array Express database, in particular, the E-MTAB-6754.processed.2.zip file to obtain the UMI counts of bone marrow-derived mononuclear phagocytes from mouse, rat, pig and rabbit. Phagocytes were either untreated (0 h) or stimulated with LPS for 2, 4 and 6 h, resulting in 12 scRNA-seq datasets per species. In addition, phagocytes from mice and rat were also treated with PIC at 2, 4 and 6 h. Notably, the dataset contains no UMI counts for PIC stimulation at 6 h for mouse 1 but has two for mouse 2 (labelled 6 and 6A). When collating the counts, the missing replicate for mouse 1 was disregarded and the PIC 6A time point—assumed to be a technical replicate—was excluded. Therefore, 20 datasets (referred as conditions herein) for the mouse, 21 datasets for the rat, 12 conditions for the pig and the rabbit dataset were considered for each gene (see Supplementary Table S4). The UMI counts were median scaled per cell using the normalize_total function of Scanpy and subsequently used for fitting mean-variance relationships and bursting modulation. Integer values, referred to as "mRNA counts" in this work were used for mathematical model fitting (see Github repository for data normalisation, UMI normalisation (Grün et al., 2014) and extraction of mRNA count distributions). Gene IDs, gene symbols and the descriptions of the genes were obtained from the Ensembl Release 103 database of the four studied species: *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Sus scrofa* (pig) and *Oryctolagus Cuniculus* (rabbit) using the BioMart web tool (Yates et al., 2020). Hagai *et al.* (2018) defined a set of 2,336 LPS-responsive genes based on differential expression in response to LPS stimulation with FDR-corrected p-value < 0.01 and existing orthologues in rabbit, rat and pig. *Il1b* and *Tnf* were added to this list—as well characterised TLR-response genes from the study of Bagnall et al. (2020)–resulting in a set of 2,338 LPS response genes with 46,740 conditions overall. Similarly, the responsive genes from the three other species were also determined. 2,586 rat genes, 1892 pig genes and 859 rabbit genes showed differential expression upon LPS stimulus. 218 one-to-one orthologue genes were found to be responsive in all species, these genes formed the analysis subset.

Fitting theoretical bursting characteristics

The sample mean (μ) and variance (σ^2) of mRNA counts were calculated for the measured mRNA count distribution for individual response genes across conditions. The mean-variance relationships ($\sigma^2 = \alpha\mu + \alpha_0$) were fitted using robust linear regression, using a Huber M-estimator with a tuning constant of 1.345, across all relevant conditions. A model's fit was considered successful if the slope (α) was statistically significant based on FDR-adjusted p-value < 0.05, and it provided a good overall fit (unweighted $R^2 > 0.6$). FDR-adjusted p-value < 0.05 was also calculated for the intercept (α_0). Assuming linear constraints of mRNA mean and variance, theoretical bursting characteristics were analytically derived, using moment estimators; burst size $b_s = \alpha_0/\mu + \alpha$, burst frequency $b_f = \mu^2/(\alpha_0+\alpha\mu)$ and $b_f = \alpha_0/(b_s(b_s-\alpha))$. Relative root mean square error, $= \sqrt{\frac{\sum_{i=1}^{N}(experimental\,data_i - model\,data_i)^2}{N\sum_{i}^{N}(model\,data_i)^2}}$, where N denoted the number of datapoints, was used to compare theoretical predictions and experimental data. Relative fold change was used to calculate the level of burst size and frequency modulation in the measured data, across all the conditions per gene:

$$burst\ size\ modulation\ per\ gene = \frac{max_{bs} - min_{bs}}{min_{bs}},$$

$$burst\ frequency\ modulation\ per\ gene = \frac{max_{bf} - min_{bf}}{min_{bf}}$$

Comparison between burst size and burst frequency modulation was quantified as the ratio of the two quantities, i.e., $modulation\ ratio = \frac{burst\ frequency\ modulation}{burst\ size\ modulation}$.

Pairwise comparison of the slopes of the mean-variance regressions

The differences in the mean-variance relationships of a gene between species were measured by pairwise comparisons between the slopes. A Student's t-test was performed to determine whether the two slopes are statistically significantly different, or not. The following formula was used to calculate the t-statistic values:

$$t_{statistic} = \frac{slope_1 - slope_2}{\sqrt{SE_{slope_1}{}^2 + SE_{slope_2}{}^2}}, \quad d.o.f. = n_1 + n_2 - 4$$

SE_{slope} represents the standard error of the value of the slope in the fitting of the robust linear regression model on the data. The degrees of freedom (d.o.f.) is dependent on the number of data points used to create the two linear regression lines compared (n_1 and n_2, respectively). p-values were determined using the cumulative distribution function of the relevant t distribution. As the four slopes were compared pairwise, six p-values were calculated per gene. p-values were corrected by the Benjamini-Hochberg procedure. Two slopes were deemed significantly different if the false discovery rate (FDR) corrected p-value was below 0.05. Subset of genes with different number of significant FDR-corrected

p-values were compared using a measure of evolutionary response divergence, such that *response divergence* = $log[1/3 \times \sum_j (log[FC\ pig] - log[FC\ glire_j])^2]$, with j =(1,2,3) corresponding to 3 glires (mouse, rat and rabbit) and FC is the fold change in response to LPS stimulation per gene (Supplementary Table S4 in (Hagai et al., 2018)).

Inference of Beta-Poisson model

Inference of Beta-Poisson model parameters (k_{on}, k_{off} and k_t) from individual scRNA-seq count distributions was performed using the profile-likelihood txburstML script (Larsson et al., 2019) downloaded from GitHub (version 1844c47be5f1ad2104cf15d425889768ec45df8b). Conditions that txburstML did not mark as "keep" (indicating convergence) were discarded. Genes with a least 10 fitted conditions per mouse (out of 20) and rat (out of 21) as well at least 6 in the pig and rabbit (out of 12) were included in the high coverage gene sets.

Modelling and inference of dynamical models of transcription

Theoretical temporal mRNA distributions for considered models of transcription were obtained using the Chemical Master Equation (CME) following our previous approach (Bagnall et al., 2020). In brief, the time evolution of the probability distribution over mRNA counts $P(\boldsymbol{X},t)$, is given by $P(\boldsymbol{X},t) = \exp[R(\theta)t]P_0(\boldsymbol{X})$, where $R(\theta)$ is a transition rate matrix describing flow of probability between different states, where a state is defined by the number of mRNA in the cell at time t and the transcriptional states of the gene's alleles. $P_0(\boldsymbol{X})$ is specified by initial data such that $\sum_X P_0(\boldsymbol{X}) = 1$. $P(\boldsymbol{X},t)$ is calculated using a fast matrix exponential function implemented in MATLAB by (Al-Mohy and Higham, 2011). All simulations begin with initial conditions of no mRNA and both gene alleles being in the '*off*' state. $R(\theta)$ depends on model structure and the parameters. In this work, we considered a *stochastic telegraph model*—with two independent alleles per gene, the activity of which switches randomly between '*off*' and '*on*' states, with the latter being permissive for mRNA transcription (Raj et al., 2006; Zenklusen et al., 2008; Suter et al., 2011; Skinner et al., 2016). The associated kinetic parameters include switching '*on*' and '*off*' rates (k_{on} and k_{off}, respectively) as well as rates of mRNA transcription and degradation (k_t and k_d, respectively). We also considered an extended model including an additional regulatory step, such that each allele exists in one of three states: an inactive '*off*', an intermediate '*I*' or an active '*on*'. Reversible stochastic transitions (with appropriate rates) occur between the inactive and intermediate (t_{on} and t_{off}), the intermediate and active states (k_{on} and k_{off}), as well as direct transition between active and inactive states (k_c). We further assume that transcription occurs only in the intermediate and active states (k_{ti} and k_t, respectively).

A genetic algorithm (GA) was implemented using the *ga* function in MATLAB and employed to estimate model parameters. We minimised an objective function given by the average absolute distance between the theoretical (CME) and measured cumulative distribution functions (CDFs) across observed mRNA counts per condition $(1/n\sum_{i=1}^{n}|CME_i - CDF_i|)$, where i's are unique mRNA counts observed in the measured distributions (for those with total unique counts n > 1). CDFs were calculated using empirical cumulative distribution function (*ecdf*). The best of 10 model fits from independent GA runs for each condition (using a population size of 100, elite count of 2, crossover factor of 0.6, 20 generations and the tournament selection function) was retained. Gene activation/inactivation rates were constrained between 0 and 1 min^{-1}, transcription was constrained between 0 and 50 mRNA counts min^{-1} per allele, which is the same order of magnitude to previous estimates (Schwanhausser et al., 2011; Suter et al., 2011; Molina et al., 2013; Skinner et al., 2016). Murine mRNA half-lives (defines as $t_{1/2} = log(2)/k_d$, where k_d is a degradation rate) were obtained from literature, when available from LPS-stimulated bone marrow derived macrophages (Hao and Baltimore, 2009; Kratochvill et al., 2011) or other cell models (Maurer et al., 1999; Raghavan et al., 2002; Park et al., 2004; Sharova et al., 2009; Kambara et al., 2014; Payne et al., 2014; Martin et al., 2017; Zainol et al., 2019). Murine half-lives were also used when fitting orthologue genes.

Akaike's Information Criterium (AIC) was used to asses model fits and perform model selection (Akaike, 1973). $AIC = 2p - 2\log[L(\Theta|X)]$ where $\log[L(\Theta|X)]$ is the log-likelihood function of the fitted mRNA count distribution given measured data X defined as $L = (\frac{(\sum_j Y_j)!}{\prod_k Y_k!}) \prod_{i=1}^{N} [P(\boldsymbol{x}_i, t)]^Y$ with Y^t being a vector of the number of cells displaying each observed state at time t (the sum of this vector is the total number of cells N), and p corresponds to number of parameters in the model; resulting in a penalty for higher complexity. Models with AIC larger than Q3+1.5(Q3-Q1), where Q1 and Q3 are the first and third quartiles of the AIC distribution per model across genes were removed to account for unsatisfactory GA fits. As a result, out of 1,507 mouse, and 1,079 orthologue (pig, rat and rabbit) conditions, 1,210 and 981 that fitted 2- and 3-state models were retained, respectively.

Statistical analyses

Statistical analysis was performed using GraphPad Prism 8 software (version 8.4.2). The D'Agostino-Pearson test was applied to test for normal (Gaussian) distribution of acquired data. Two-sample comparison was conducted using non-parametric Mann Whitney test. For analyses of variance Kruskal-Wallis ANOVA with Dunn's multiple comparisons test was performed. Coefficient of determination (R^2) was used to assess regression fits; Spearman correlation coefficient r was used to test association between other variables.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/ppaszek/TLR_bursting; https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-6754.

Author contributions

NA performed analyses presented in the manuscript. DN and ZW performed preliminary analyses and developed Python codes.

MM and PP provided supervision and conceptualisation. PP with assistance of NA and MM wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

NA was supported by Wellcome Trust PhD Studentship. This work was also supported by BBSRC (BB/R007691/1).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1176107/full#supplementary-material

References

Adamson, A., Boddington, C., Downton, P., Rowe, W., Bagnall, J., Lam, C., et al. (2016). Signal transduction controls heterogeneous NF-κB dynamics and target gene expression through cytokine-specific refractory states. *Nat. Commun.* 7, 12057. doi:10.1038/ncomms12057

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu Akaike* (Berlin, Germany: Springer).

Al-Mohy, A. H., and Higham, N. J. (2011). Computing the action of the matrix exponential, with an application to exponential integrators. *Siam J. Sci. Comput.* 33, 488–511. doi:10.1137/100788860

Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H. B., Trombetta, J. J., et al. (2015). Pathogen cell-to-cell variability drives heterogeneity in host immune responses. *Cell* 162, 1309–1321. doi:10.1016/j.cell.2015.08.027

Bagnall, J., Boddington, C., England, H., Brignall, R., Downton, P., Alsoufi, Z., et al. (2018). Quantitative analysis of competitive cytokine signaling predicts tissue thresholds for the propagation of macrophage activation. *Sci. Signal* 11, eaaf3998. doi:10.1126/scisignal.aaf3998

Bagnall, J., Rowe, W., Alachkar, N., Roberts, J., England, H., Clark, C., et al. (2020). Gene-specific linear trends constrain transcriptional variability of the toll-like receptor signaling. *Cell Syst.* 11, 300–314. doi:10.1016/j.cels.2020.08.007

Bass, V. L., Wong, V. C., Bullock, M. E., Gaudet, S., and Miller-Jensen, K. (2021). TNF stimulation primarily modulates transcriptional burst size of NF-κB-regulated genes. *Mol. Syst. Biol.* 17, e10127. doi:10.15252/msb.202010127

Brennan, J. J., and Gilmore, T. D. (2018). Evolutionary origins of toll-like receptor signaling. *Mol. Biol. Evol.* 35, 1576–1587. doi:10.1093/molbev/msy050

Bryant, C. E., Symmons, M., and Gay, N. J. (2015). Toll-like receptor signalling through macromolecular protein complexes. *Mol. Immunol.* 63, 162–165. doi:10.1016/j.molimm.2014.06.033

Buckley, J. M., Wang, J. H., and Redmond, H. P. (2006). Cellular reprogramming by gram-positive bacterial components: A review. *J. Leukoc. Biol.* 80, 731–741. doi:10.1189/jlb.0506312

Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V., McCollum, J. M., Cox, C. D., et al. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17454–17459. doi:10.1073/pnas.1213530109

Dar, R. D., Shaffer, S. M., Singh, A., Razooky, B. S., Simpson, M. L., Raj, A., et al. (2016). Transcriptional bursting explains the noise-versus-mean relationship in mRNA and protein levels. *Plos One* 11, e0158298. doi:10.1371/journal.pone.0158298

de la Higuera, L., Lopez-Garcia, M., Castro, M., Abourashchi, N., Lythe, G., and Molina-Paris, C. (2019). Fate of a naive T cell: A stochastic journey. *Front. Immunol.* 10, 194. doi:10.3389/fimmu.2019.00194

Dey, S. S., Foley, J. E., Limsirichai, P., Schaffer, D. V., and Arkin, A. P. (2015). Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Mol. Syst. Biol.* 11, 806. doi:10.15252/msb.20145704

Einarsson, H., Salvatore, M., Vaagenso, C., Alcaraz, N., Lange, J., Rennie, S., et al. (2022). Promoter sequence and architecture determine expression variability and confer robustness to genetic variants. *eLife* 11, e80943. doi:10.7554/eLife.80943

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186. doi:10.1126/science.1070919

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer control of transcriptional bursting. *Cell* 166, 358–368. doi:10.1016/j.cell.2016.05.025

Gay, N. J., Symmons, M. F., Gangloff, M., and Bryant, C. E. (2014). Assembly and localization of Toll-like receptor signalling complexes. *Nat. Rev. Immunol.* 14, 546–558. doi:10.1038/nri3713

Gomez-Schiavon, M., Chen, L. F., West, A. E., and Buchler, N. E. (2017). BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. *Genome Biol.* 18, 164. doi:10.1186/s13059-017-1297-9

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* 11, 637–640. doi:10.1038/nmeth.2930

Hagai, T., Chen, X., Miragaia, R. J., Rostom, R., Gomes, T., Kunowska, N., et al. (2018). Gene expression variability across cells and species shapes innate immunity. *Nature* 563, 197–202. doi:10.1038/s41586-018-0657-2

Hao, S., and Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat. Immunol.* 10, 281–288. doi:10.1038/ni.1699

Harper, C. V., Finkenstadt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., et al. (2011). Dynamic analysis of stochastic transcription cycles. *PLoS Biol.* 9, e1000607. doi:10.1371/journal.pbio.1000607

Haygood, R., Babbitt, C. C., Fedrigo, O., and Wray, G. A. (2010). Contrasts between adaptive coding and noncoding changes during human evolution. *Proc. Natl. Acad. Sci.* 107, 7853–7857. doi:10.1073/pnas.0911249107

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment, Computing in Science and Engineering. *IEEE Computer Society* 9 (3), 99–104. doi:10.1109/MCSE.2007.55

Hoppe, C., Bowles, J. R., Minchington, T. G., Sutcliffe, C., Upadhyai, P., Rattray, M., et al. (2020). Modulation of the promoter activation rate dictates the transcriptional response to graded BMP signaling levels in the Drosophila embryo. *Dev. Cell* 54, 727–741. doi:10.1016/j.devcel.2020.07.007

Iakovlev, M., Faravelli, S., and Becskei, A. (2021). Gene families with stochastic exclusive gene choice underlie cell adhesion in mammalian cells. *Front. Cell Dev. Biol.* 9, 642212. doi:10.3389/fcell.2021.642212

Kalliara, E., Kardynska, M., Bagnall, J., Spiller, D. G., Müller, W., Ruckerl, D., et al. (2022). Post-transcriptional regulatory feedback encodes JAK-STAT signal memory of interferon stimulation. *Front. Immunol.* 13, 947213. doi:10.3389/fimmu.2022.947213

Kambara, H., Niazi, F., Kostadinova, L., Moonka, D. K., Siegel, C. T., Post, A. B., et al. (2014). Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res.* 42, 10668–10680. doi:10.1093/nar/gku713

Kratochvill, F., Machacek, C., Vogl, C., Ebner, F., Sedlyarov, V., Gruber, A. R., et al. (2011). Tristetraprolin-driven regulatory circuit controls quality and timing of mRNA decay in inflammation. *Mol. Syst. Biol.* 7, 560. doi:10.1038/msb.2011.93

Larson, D. R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D. S., and Singer, R. H. (2013). Direct observation of frequency modulated transcription in single cells using light activation. *Elife* 2, e00750. doi:10.7554/eLife.00750

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., et al. (2019). Genomic encoding of transcriptional burst kinetics. *Nature* 565, 251–254. doi:10.1038/s41586-018-0836-1

Lee, R. E., Walker, S. R., Savery, K., Frank, D. A., and Gaudet, S. (2014). Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. *Mol. Cell* 53, 867–879. doi:10.1016/j.molcel.2014.01.026

Lind, M. I., and Spagopoulou, F. (2018). Evolutionary consequences of epigenetic inheritance. *Heredity* 121, 205–209. doi:10.1038/s41437-018-0113-y

Lu, Y., Xue, Q., Eisele, M. R., Sulistijo, E. S., Brower, K., Han, L., et al. (2015). Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *Proc. Natl. Acad. Sci. U. S. A.* 112, E607–E615. doi:10.1073/pnas.1416756112

Luecken, M. D., and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* 15, e8746. doi:10.15252/msb.20188746

Luo, X., Qin, F., Xiao, F., and Cai, G. (2022). BISC: Accurate inference of transcriptional bursting kinetics from single-cell transcriptomic data. *Briefings Bioinforma.* 23, bbac464. doi:10.1093/bib/bbac464

Martin, L. J., Smith, S. B., Khoutorsky, A., Magnussen, C. A., Samoshkin, A., Sorge, R. E., et al. (2017). Epiregulin and EGFR interactions are involved in pain processing. *J. Clin. Invest.* 127, 3353–3366. doi:10.1172/JCI87406

Maurer, F., Tierney, M., and Medcalf, R. L. (1999). An AU-rich sequence in the 3'-UTR of plasminogen activator inhibitor type 2 (PAI-2) mRNA promotes PAI-2 mRNA decay and provides a binding site for nuclear HuR. *Nucleic Acids Res.* 27, 1664–1673. doi:10.1093/nar/27.7.1664

Megaridis, M. R., Lu, Y. Y., Tevonian, E. N., Junger, K. M., Moy, J. M., Bohn-Wippert, K., et al. (2018). Fine-tuning of noise in gene expression with nucleosome remodeling. *Apl. Bioeng.* 2, 026106. doi:10.1063/1.5021183

Molina, N., Suter, D. M., Cannavo, R., Zoller, B., Gotic, I., and Naef, F. (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20563–20568. doi:10.1073/pnas.1312310110

Morris, M. C., Gilliam, E. A., and Li, L. (2014). Innate immune programing by endotoxin and its pathological consequences. *Front. Immunol.* 5, 680. doi:10.3389/fimmu.2014.00680

Nicolas, D., Phillips, N. E., and Naef, F. (2017). What shapes eukaryotic transcriptional bursting? *Mol. Biosyst.* 13, 1280–1290. doi:10.1039/c7mb00154a

Nicolas, D., Zoller, B., Suter, D. M., and Naef, F. (2018). Modulation of transcriptional burst frequency by histone acetylation. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7153–7158. doi:10.1073/pnas.1722330115

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., et al. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLOS Biol.* 3, e170. doi:10.1371/journal.pbio.0030170

Nomiyama, H., Osada, N., and Yoshie, O. (2010). The evolution of mammalian chemokine genes. *Cytokine Growth Factor Rev.* 21, 253–262. doi:10.1016/j.cytogfr.2010.03.004

Nourmohammad, A., Rambeau, J., Held, T., Kovacova, V., Berg, J., and Lässig, M. (2017). Adaptive evolution of gene expression in Drosophila. *Cell Rep.* 20, 1385–1395. doi:10.1016/j.celrep.2017.07.033

Ochiai, H., Hayashi, T., Umeda, M., Yoshimura, M., Harada, A., Shimizu, Y., et al. (2020). Genome-wide kinetic properties of transcriptional bursting in mouse embryonic stem cells. *Sci. Adv.* 6, eaaz6699. doi:10.1126/sciadv.aaz6699

Park, J. H., Yu, Q., Erman, B., Appelbaum, J. S., Montoya-Durango, D., Grimes, H. L., et al. (2004). Suppression of IL7Ralpha transcription by IL-7 and other prosurvival cytokines: A novel mechanism for maximizing IL-7-dependent T cell survival. *Immunity* 21, 289–302. doi:10.1016/j.immuni.2004.07.016

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature* 427, 415–418. doi:10.1038/nature02257

Payne, T. L., Blackinton, J., Frisbee, A., Pickeral, J., Sawant, S., Vandergrift, N. A., et al. (2014). Transcriptional and posttranscriptional regulation of cytokine gene expression in HIV-1 antigen-specific $CD8^+$ T cells that mediate virus inhibition. *J. Virology* 88, 9514–9528. doi:10.1128/JVI.00802-14

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12, 2825–2830.

Phillips, N. E., Mandic, A., Omidi, S., Naef, F., and Suter, D. M. (2019). Memory and relatedness of transcriptional activity in mammalian cell lineages. *Nat. Commun.* 10, 1208. doi:10.1038/s41467-019-09189-8

Raghavan, A., Ogilvie, R. L., Reilly, C., Abelson, M. L., Raghavan, S., Vasdewani, J., et al. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Res.* 30, 5529–5538. doi:10.1093/nar/gkf682

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4, e309. doi:10.1371/journal.pbio.0040309

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* 5, 877–879. doi:10.1038/nmeth.1253

Reback, J., McKinney, W., Jbrockmendel Van Den Bossche, J., Augspurger, T., Cloud, P., et al. (2020). pandas-dev/pandas: Pandas 1.0.5. doi:10.5281/ZENODO.3898987

Robles-Rebollo, I., Cuartero, S., Canellas-Socias, A., Wells, S., Karimi, M. M., Mereu, E., et al. (2022). Cohesin couples transcriptional bursting probabilities of inducible enhancers and promoters. *Nat. Commun.* 13, 4342. doi:10.1038/s41467-022-31192-9

Rybakova, K. N., Bruggeman, F. J., Tomaszewska, A., Moné, M. J., Carlberg, C., and Westerhoff, H. V. (2015). Multiplex eukaryotic transcription (In)activation: Timing, bursting and cycling of a ratchet clock mechanism. *PLoS Comput. Biol.* 11, e1004236. doi:10.1371/journal.pcbi.1004236

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi:10.1038/nature10098

Seabold, S., and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python, *9th Python in Science Conference*.

Shaffer, S. M., Emert, B. L., Reyes Hueros, R. A., Cote, C., Harmange, G., Schaff, D. L., et al. (2020). Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell* 182, 947–959. doi:10.1016/j.cell.2020.07.003

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240. doi:10.1038/nature12172

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510, 363–369. doi:10.1038/nature13437

Shand, L., and Li, B. (2017). Modeling nonstationarity in space and time. *Biometrics* 73, 759–768. doi:10.1111/biom.12656

Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* 16, 45–58. doi:10.1093/dnares/dsn030

Skinner, S. O., Xu, H., Nagarkar-Jaiswal, S., Freire, P. R., Zwaka, T. P., and Golding, I. (2016). Single-cell analysis of transcription kinetics across the cell cycle. *Elife* 5, e12175. doi:10.7554/eLife.12175

Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M., and Sorger, P. K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459, 428–432. doi:10.1038/nature08012

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, 472–474. doi:10.1126/science.1198817

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., et al. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538. doi:10.1126/science.1188308

Van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering* 13 (2), 22–30. doi:10.1109/MCSE.2011.37

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17 (3), 261–272. doi:10.1038/s41592-019-0686-2

Vu, T. N., Wills, Q. F., Kalari, K. R., Niu, N., Wang, L., Rantalainen, M., et al. (2016). Beta-Poisson model for single-cell RNA-seq data analyses. *Bioinformatics* 32, 2128–2135. doi:10.1093/bioinformatics/btw202

Wang, Y., Qi, J., Shao, J., and Tang, X. Q. (2020). Signaling mechanism of transcriptional bursting: A technical resolution-independent study. *Biol. (Basel)* 9, 339. doi:10.3390/biology9100339

Wang, Z., Guo, Y., and Gong, H. (2018). An integrative analysis of time-varying regulatory networks from high-dimensional data. *Proc. IEEE Int. Conf. Big Data* 2018, 3798–3807. doi:10.1109/BigData.2018.8622361

Waskom, M., Botvinnik, O., Ostblom, J., Gelbart, M., Lukauskas, S., Hobson, P., et al. (2020). mwaskom/seaborn: v0.10.1 (April 2020). doi:10.5281/ZENODO.3767070

Wolf, F. A., Angerer, P., and Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19 (1), 15. doi:10.1186/s13059-017-1382-0

Wong, V. C., Bass, V. L., Bullock, M. E., Chavali, A. K., Lee, R. E. C., Mothes, W., et al. (2018). NF-κB-Chromatin interactions drive diverse phenotypes by modulating transcriptional noise. *Cell Rep.* 22, 585–599. doi:10.1016/j.celrep.2017.12.080

Xue, Q., Lu, Y., Eisele, M. R., Sulistijo, E. S., Khan, N., Fan, R., et al. (2015). Analysis of single-cell cytokine secretion reveals a role for paracrine signaling in coordinating macrophage responses to TLR4 stimulation. *Sci. Signal.* 8, ra59. doi:10.1126/scisignal.aaa2155

Yang, X., Wang, Z., Wu, Y., Zhou, T., and Zhang, J. (2022). Kinetic characteristics of transcriptional bursting in a complex gene model with cyclic promoter structure. *Math. Biosci. Eng.* 19, 3313–3336. doi:10.3934/mbe.2022153

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., et al. (2020). Ensembl 2020. *Nucleic Acids Res.* 48 (D1), D682–D688. doi:10.1093/nar/gkz966

Zainol, M. I. B., Kawasaki, T., Monwan, W., Murase, M., Sueyoshi, T., and Kawai, T. (2019). Innate immune responses through Toll-like receptor 3 require human-antigen-R-mediated Atp6v0d2 mRNA stabilization. *Sci. Rep.* 9, 20406. doi:10.1038/s41598-019-56914-w

Zenklusen, D., Larson, D. R., and Singer, R. H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15, 1263–1271. doi:10.1038/nsmb.1514

Zoller, B., Little, S. C., and Gregor, T. (2018). Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting. *Cell* 175, 835–847. doi:10.1016/j.cell.2018.09.056

Zoller, B., Nicolas, D., Molina, N., and Naef, F. (2015). Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* 11, 823. doi:10.15252/msb.20156257

Supplementary Figures and Tables

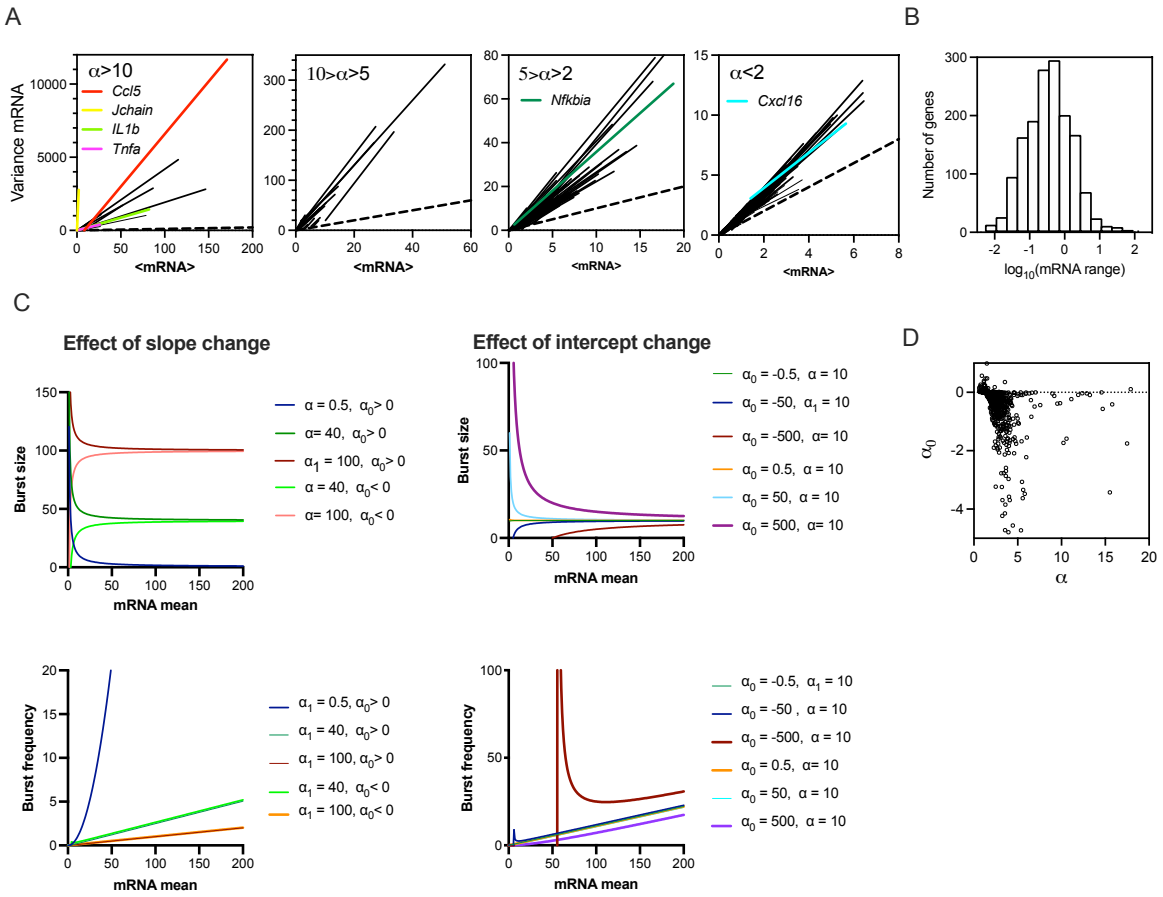


Figure S1. Analysis of the variability in the TLR responses. A. Fitted regression lines for the 1,551 high confidence genes, shown are genes with different range of the slope α. Highlighted in different colours are fits for the individual genes. Broken line indicates $\mu=\sigma^2$ line. **B**. Histogram of the measured mRNA response range for the 1,551 high confidence genes. **C**. Effect of the slope (left) and intercept (right) of the mean-variance relationship on the burst size and burst frequency modulation. Shown are simulated burst size and frequency modulation schemes for a range of α and α_0 (as indicated on the graph). **D**. Modulation schemes for *Jchain* gene. Shown is the comparison between theoretical relationships based of fitted mean-variance relationships (in red) and corresponding estimates from data (open circles). Equation for fitted mean-variance relationships highlighted in the top left panel, respectively. **E**. Relationship between the slope (α) and in the intercept (α_0) across fitted 1,551 high confidence genes.

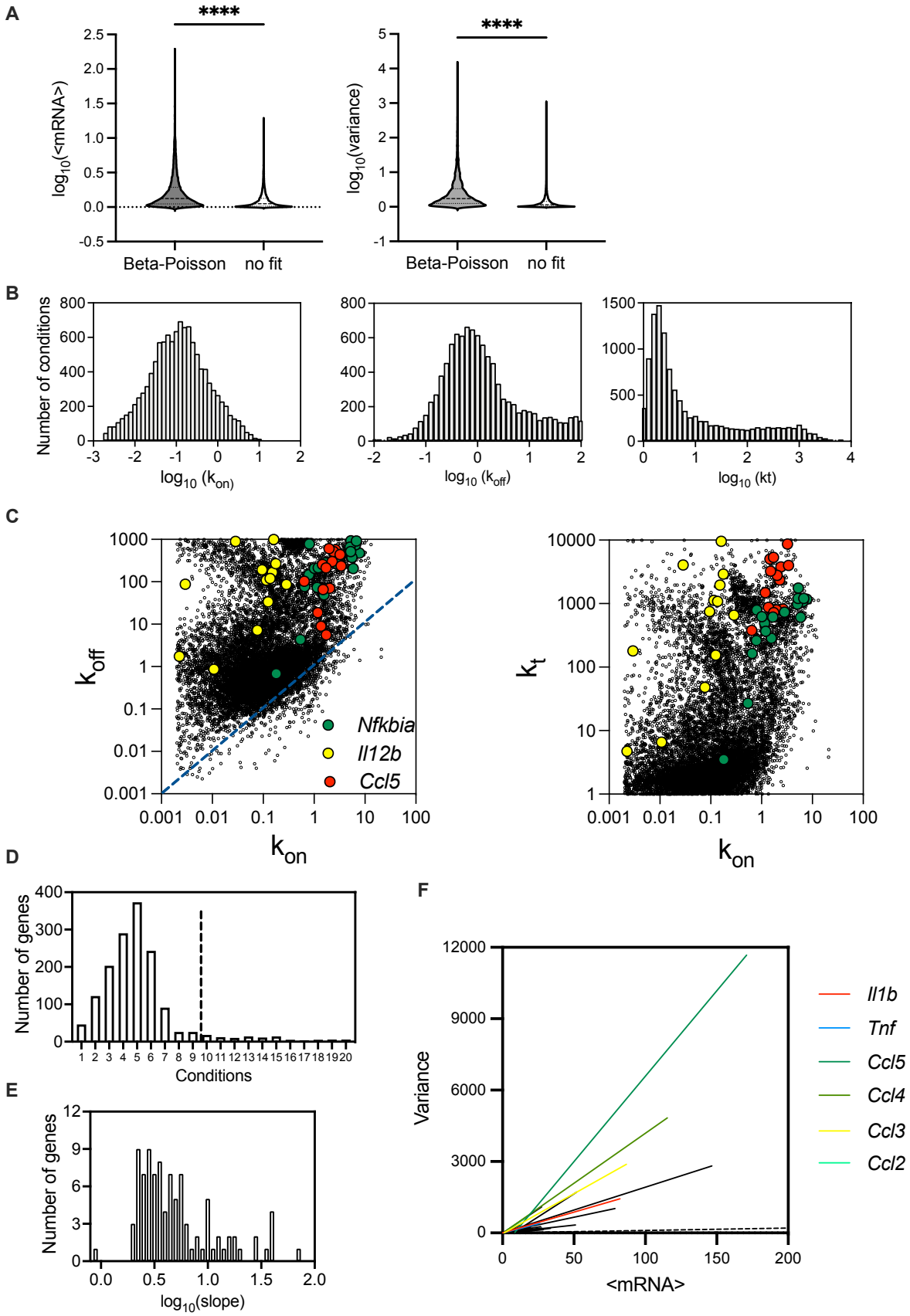

A
****
****
log10(<mRNA>)
log10(variance)
Beta-Poisson
no fit
Beta-Poisson
no fit
B
Number of conditions
log10 (kon)
log10 (koff)
log10 (kt)
C
koff
kon
Nfkbia
Il12b
Ccl5
kt
kon
D
Number of genes
Conditions
E
Number of genes
log10(slope)
F
Variance
<mRNA>
Il1b
Tnf
Ccl5
Ccl4
Ccl3
Ccl2

Figure S2. Inferred kinetic parameter rates for two-state telegraph model using Beta-Poisson model. A. Comparison between the 1551 high confidence genes across all conditions that either fit or do not fit the Beta-Poisson model. **B.** Histogram of fitted k_{on}, k_{off} and k_t across 7704 conditions for 1,519 high confidence genes. Inference performed using profile likelihood of the Beta-Poisson model. Parameters units are expressed per degradation half-life **C**. Relationship between inferred k_{on} vs. k_{off} rates (left) and k_{on} vs. k_t (right) across parameters from A. Rates for *Il12*, *Nfkbia* and *Ccl5* highlighted in different colours. Identity line depicted with a broken line. **D.** Histogram of the number of inferred conditions across 1,159 high confidence genes. Broken line highlights the threshold for at least 10 conditions fitted per gene. **E.** Histogram of the fitted regression slopes for the 96 high coverage gene set. **F.** Fitted regression lines for the 96 high coverage genes. Highlighted in colour are fits for the individual genes of interest. Broken line indicates $\mu=\sigma^2$ line.

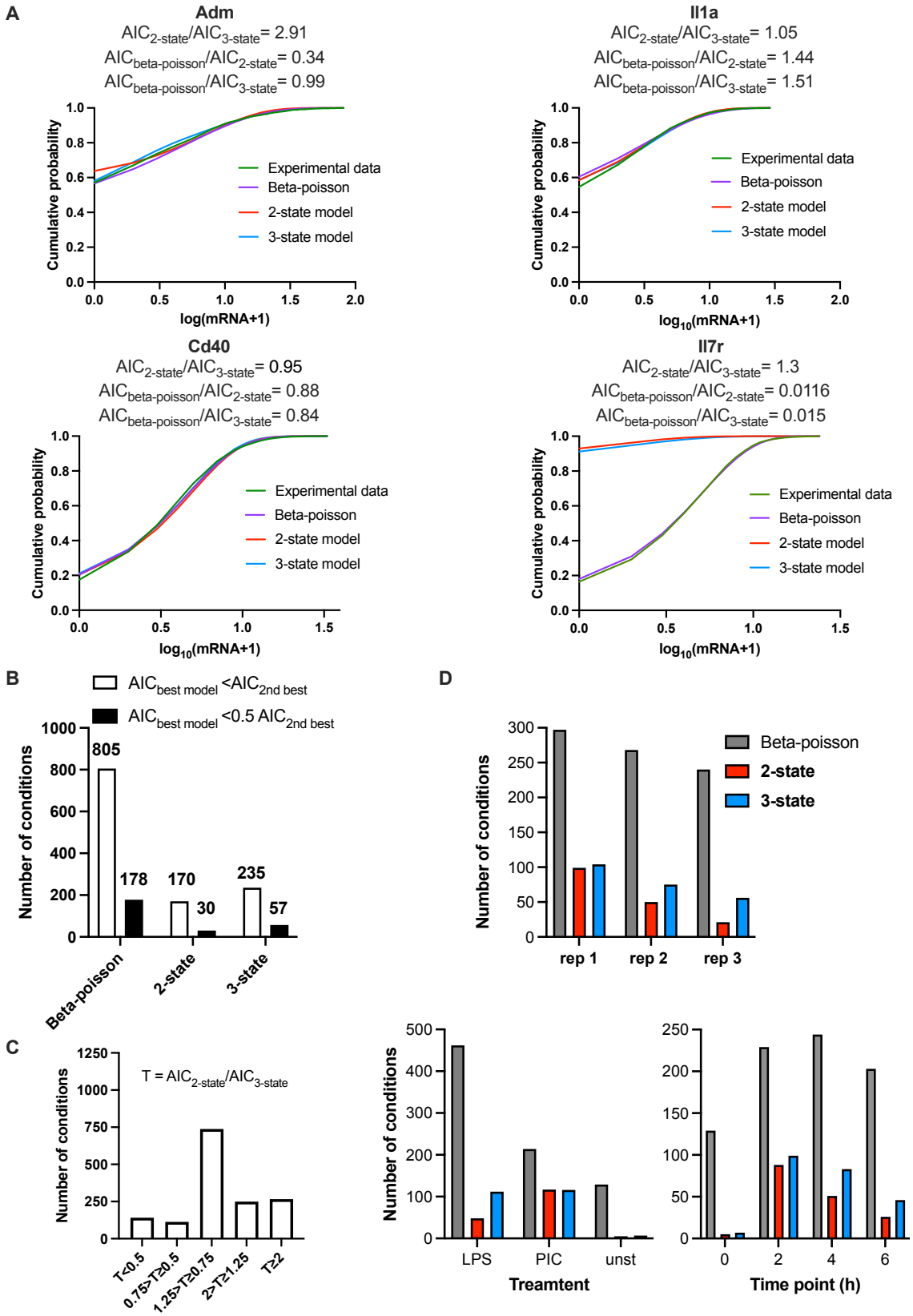


Figure S3. Analysis of stochastic models of transcription. A. Comparison between the fitted and measured scRNA-seq count distributions for few gene examples. Shown are cumulative probability distribution of data (in green) vs. the corresponding Beta-Poisson, 2-state and 3-

state model fits (in blue, red and violet, respectively) for *Adm* (LPS, 2h, replicate 1), *Il1α* (PIC, 2h, replicate 1), *Cd40* (LPS, 4h, replicate 1) and *Il7r* (0h, replicate 2) genes. Ratios of respective AICs between models highlighted on top. **B**. Summary of comparing Beta-Poisson, 2-state and 3-state model fits across the high coverage genes and conditions. Best models defined either by AIC smaller (in white) or 2-fold smaller (in black) than the next best model. **C**. Summary of 2- and 3-state model fits across a range of thresholds $T = AIC_{2\text{-}state}/AIC_{3\text{-}state}$ for the fitted 96 high coverage genes across all conditions. **D.** Relationships between the number of Beta-Poisson, 2-state and 3-state model fits for the 96 high coverage genes across all conditions. Best fit model defined by $AIC_{best\ model} < AIC_{2nd\ best}$.

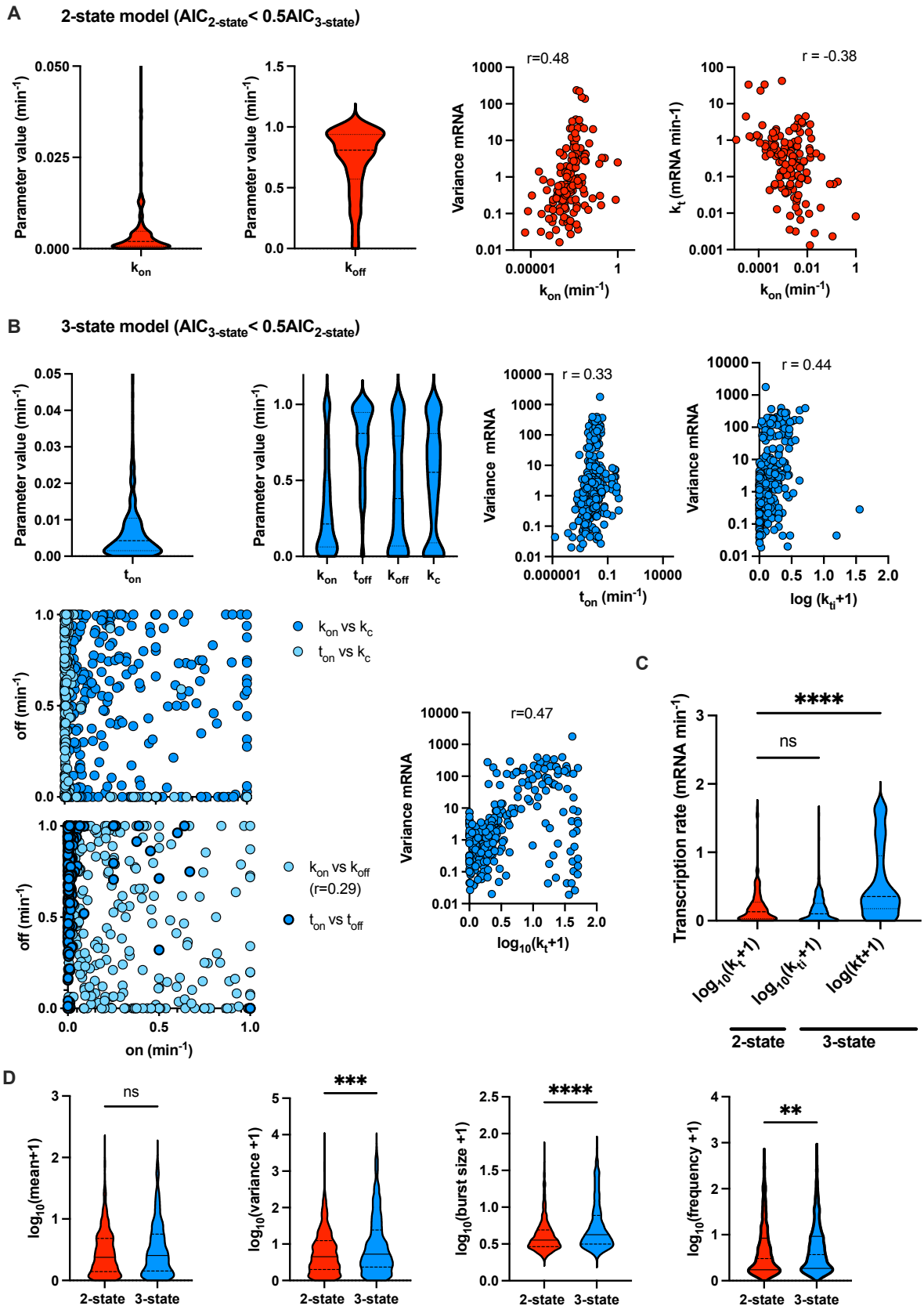


Figure S4. Model-based analysis of transcriptional bursting. **A**. Summary of 2-state model fits defined for 173 conditions such that $AIC_{2\text{-}state}<0.5AIC_{3\text{-}state}$ (as in Fig. 4C). Shown is the distribution of fitted k_{on} (min^{-1}) and k_{off} (min^{-1}) rates as well as Spearman correlation coefficient

r with mRNA variance. **B**. Summary of 3-state model fits defined for 275 conditions such that $AIC_{3\text{-}state}<0.5AIC_{2\text{-}state}$ (as in Fig. 4B). Shown is the distribution of fitted rates as well as Spearman correlation coefficient r with mRNA variance (and between selected rates). **C**. Comparison between fitted transcription rates for 2-state and 3-state models (as in A and B, respectively). Statistical significance assessed with Kruskall-Wallis test with Dunn's correction for multiple comparisons (* p value<0.05, *** p value <0.001). **D**. Analysis of transcriptional bursting across high coverage genes and conditions fitted by 2-state vs 3-state models. Shown is the comparison between best fit 2- and 3-state models in terms of mean mRNA expression, variance, burst size and frequency. Best fit defined by $AIC_{best\ model}<AIC_{2nd\ best}$ (from Fig. S3C). Burst size and frequency calculated per condition using moment estimators. Statistical significance assessed with Mann-Whitney test (* p value<0.05, **** p value <0.0001, ns- not significant).

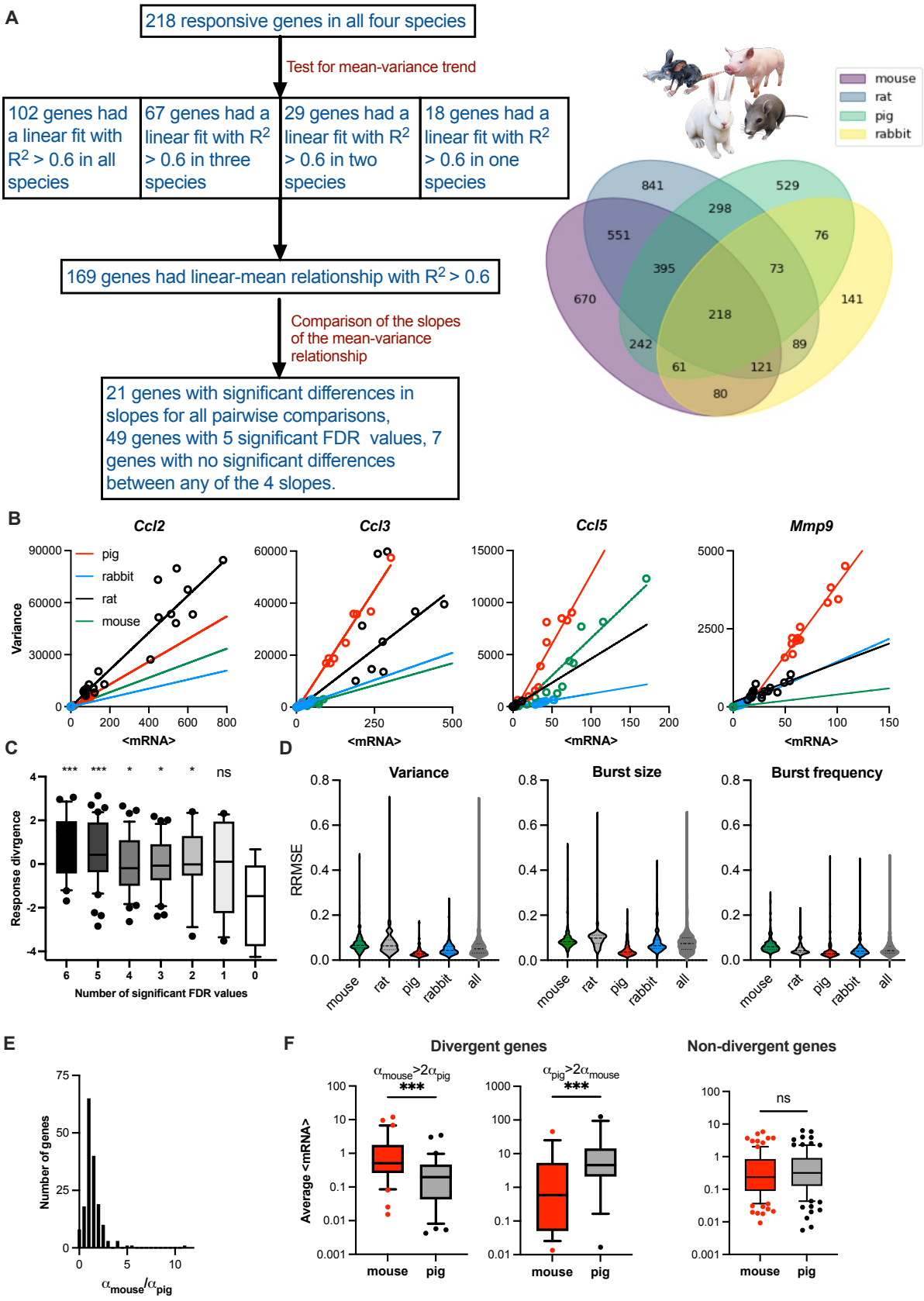


Figure S5. Analysis of transcriptional bursting across species. A. Schematic diagram of data analysis; 169 orthologue genes exhibiting good mean-variance fits ($R^2>0.6$) statistically tested for differences in the slope of the linear fit. Right: Venn diagram of TLR response

orthologue genes in at least one of the species studied by Hagai et al. (2018). **B**. Fitted mean-variance relationships for a subset of orthologue genes across species. Shown is the comparison between the fitted mean-variance relationships (in solid lines, colour-coded by species) and corresponding data (circles). **C**. Evolutionary response divergence across orthologue gene subsets defined by the number of statistically significant FDRs between fitted regression slopes across four species (as in Table S2). Statistical significance assessed using ordinary ANOVA with Dunnett's correction for multiple comparisons (*** p-val<0.001, * pval<0.05, ns – not significant). **D**. Global modulation of transcriptional busting across species. Shown is the comparison between fitted mean-variance relationship and theoretical burst size and frequency modulation schemes vs. relationships derived from data. Shown is a violin plot of relative root mean square error (RRMSE) of 169 orthologue genes. **E**. Histogram of the slope ratio ($\alpha_{mouse}/\alpha_{pig}$) for the 169 orthologue genes between mouse and pig. α_{mouse} and α_{pig} denote slopes of the fitted mean-variance relationships for each pair of species per gene. **F**. Analysis of divergent and non-divergent mouse and pig TLR-response genes. Shown are box plots of average mRNA expression per gene stratified into divergent ($\alpha_{mouse} > 2\alpha_{pig}$ or $\alpha_{pig} > 2\alpha_{pig}$) and complementary non-divergent group (31, 15 and 123 orthologue genes, respectively). Statistical significance assessed with a paired Wilcoxon test (**** p-val<0.0001, *** p-val <0.001, ns not significant).

Table S1. Fitted mean-variance relationships for the mouse TRL response genes.

Table S2. Modulation of transcriptional bursting across 1,551 mouse high confidence genes.

Table S3. Modelling of scRNA-seq count distributions.

Species	Number of cells	Total number of genes	Genes showing expression	Number of conditions
Mouse	53086	22048	16798	20
Rat	50185	22277	16780	21
Pig	23469	21607	15602	12
Rabbit	34528	19293	14480	12

Table S4. Number of phagocyte cells and genes measured in each single cell in the four species. Only the genes showing expression under at least one condition were studied

Number of significant FDR values	Genes
6	*Car4, Ccl2, Ccl4, Ccl5, Cxcl10, Ehd1, F3, Ier3, Ifit2, Ifnb1, Inhba, N4bp1, Nampt, Nlrp3, Parp9, Sema3c, Serpinb2, Slamf7, Tagln2, Tnfaip3, Tnfsf15*
5	*Adora2a, Arrdc3, Atad1, Cblb, Ccl20, Ccl3, Ccrl2, Cflar, Cmpk2, Csrnp1, Fam105a, Hmgcs1, Ifi44, Il10, Il1a, Il27, Il4ra, Irf1, Mef2c, Mmp3, Mmp9, Mxd1, Nabp1, Nfkbiz, Nrp2, Nub1, Parp11, Pik3ap1, Pim1, Rab32, Rasgef1b, Rel, Rnd1, Rnf19a, Sdc4, Serpinb8, Slamf1, Slc46a3, Snx10, Socs1, Srgn, Stat3, Tcf7l2, Tfec, Tnfaip6, Tnfsf10, Ttc39b, Txnip, Zc3hav1*
4	*A230050P20Rik, Acsl1, Cd274, Cd40, Cd53, Cdkn2c, Cxcl9, Cxcr4, Dusp2, Fas, Fgd4, Fmr1, Lpxn, Manf, Marcks, Mov10, Nbr1, Nr3c1, Olr1, Plekho1, Ppa1, Ppp1r15a, Psma6, Rgs10, Samsn1, Slc17a5, Slc29a3, Slc37a2, Tiparp, Tnip1, Tra2a, Trim25, Ulk1, Vcan, Ypel3*
3	*Amacr, Arl5b, Atp10a, Birc3, Ccng2, Coprs, Gmnn, Hbegf, Hhex, Icam1, Jak2, Mafb, Mb21d1, Mical1, Mxd4, Nfkbia, Nmi, Npc1, Nr1d2, Nr4a3, Pcgf5, Plk2, Pnrc1, Rnd3, Rnf19b, Sh3pxd2b, Smarca2, Tdrd7, Tfdp2, Traf3ip2, Trim26, Uap1, Wars, Xrn1*
2	*Baz1a, Ccdc34, Gmpr, Nfkb2, Nfkbib, Plekhm1, Prkag2, Rybp, Skil, Tmem51, Uqcc3, Xpc*
1	*Csrp2, Fam98c, Frmd4b, Gtf2i, Ldlrap1, Lpar6, Mapk6, Rasa2, Rragd, St6gal1, Top1*
0	*Arhgap4, Camk2g, Cbx8, Crot, Hdac5, Tmbim6, Uri1*

Table S5. Pairwise comparison of the slopes of the mean-variance regression lines was performed between each two species. The table shows the number of significant FDR values (<0.05) obtained for each of the 169 orthologue genes studied.

Table S6. Analysis of TLR response variability across species.

Chapter 4 Transcriptional variability as a heritable trait

In this chapter, I explore whether cells can retain a memory of their transcriptional state through cell division, leading to persistent variations in gene expression over time across several generations. Heritability measure of a gene is defined by the variability between clonal population compared to parental population variability, with the coefficient of variation used as the measure of population variability.

Through analysis of bulk and single-cell RNA-seq data across clonal macrophage populations, I reveal heritable TLR-dependent and -independent genes and explore the differences in their transcriptional regulation across clones. The fundamental question behind this work is whether our treatment of gene expression as a stochastic process is accurate or whether we need to attune the commonly used models to consider some deterministic aspects of the process in order to formulate more realistic models of transcriptional activity.

4.1 Journal paper: Single-cell gene expression patterns of the Toll-like receptor signalling are heritable traits

Article

Single-cell gene expression patterns of the Toll-like receptor signalling are heritable traits

Nissrin Alachkar[1, †], Josephine Moran[1, †], Mark Muldoon[2], Ian Roberts[1], Abhyduai Singh[3], and Pawel Paszek[1, *]

[1]Division of Immunology, Immunity to Infection and Respiratory Medicine, Lydia Becker Institute of Immunology and Inflammation, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom, [2]Department of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, UK, [3]Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA, [*]Correspondence: Tel: +44 (0) 161 275 1743, email: pawel.paszek@manchester.ac.uk, [†]These authors contributed equally to this work

Abstract

Activation of innate immunity at the single-cell level is a heterogenous process; however the origins of this variability - fundamentally linked to the control of immune responses - remain unknown. Here we combine classical fluctuation tests with genomic approaches to understand heritability of single-cell gene expression patterns in the evolutionarily conserved toll like receptor (TLR) system. Using population-level RNA-seq, we measured variability between clonal populations of immortalised Bone Marrow Derived Macrophages and showed that over multiple cell divisions approximately 7% of TLR4-dependent (and 15% of upregulated) genes, including immune cytokine and effector genes, exhibited transcriptional heritability. In contrast, 2% of TLR4-independent genes exhibited heritability, albeit with a higher level, suggesting different timescales of transcriptional memory. To validate population-level analyses and better understand inter-clone differences, we used single-cell RNA sequencing (scRNA-seq) and assayed mRNA distributions of clonal and parental populations to TLR4 stimulation. We found that individual heritable genes maintained response heterogeneity, which mathematically can be attributed to frequency modulation of transcriptional bursting across clonal populations. Overall, we demonstrate a prevalent long-term TLR-mediated transcriptional heritability and provide a theoretical basis for control of variability in heritable transcriptional traits.

Introduction

In mammals the cellular defence system against foreign threats involves the evolutionarily conserved toll-like receptor (TRL) system [1]. Paradoxically, the activation of TLR signalling is inherently a heterogenous process with target genes exhibiting a substantial transcriptional variability at the single cell level [2]. As a consequence, specific effector molecules like Tumour Necrosis Factor α (TNFα), Interleukin 1β (IL1β) or type I interferons (IFN-I) are produced by small (and often non-overlapping) subsets of genetically identical innate immune cells [3–6]. The response heterogeneity of TLR signalling is evolutionarily conserved across species [7] suggesting an important and beneficial role during immune response of inflammation [8]. However, the fundamental question regarding the nature of this heterogeneity - whether single cell innate immune responses are stochastic or predetermined - has not been fully elucidated.

The single-cell TLR-mediated transcriptional response is thought to reflect cell-intrinsic and extrinsic signalling events, including activation of key transcription factors such as Nuclear Factor κB (NF-κB) [9–13], as well as paracrine signalling, epigenetic regulation and the biology of the pathogen, among many factors [3, 5, 14, 15]. Two genetically identical cells may behave differently because they could be in different states [16–19] or be subject to stochastic fluctuations in their local environment [20, 21]. Transcriptional noise has been predominately attributed to the process of transcriptional bursting, i.e., random changes of gene activity resulting in stochastic transcription events across different cell types and tissues [22–28], including in the context of immune signalling [4, 15, 29–31]. However, recent analyses demonstrate that non-genetic heterogeneity, such as responses of rare precocious cells, persist for generations in dividing populations, thus revealing long-term transcriptional memory [16, 32, 33]. In the context of the TLR system, NF-κB signalling has been shown to exhibit correlated dynamical responses not only over one or few cell divisions [17, 34] but over days and months in clonally-derived populations [35]. Our previous data also demonstrates that biphasic TRL4-dependent expression of IL1β and IL1α, located in the same gene cluster, is highly correlated and subject to epigenetic control [4]. Furthermore, in the TLR system, heritable fates determine the early IFN-I responses [6], while heritable receptor expression regulates “all-or-nothing” activation patterns [34]. This suggests that transcriptional heritability plays important roles in determining the overall heterogeneity of the immune cell population responding to TLR stimulation.

In order to distinguish between heritable and non-heritable variability in TLR-dependent transcription, we used MemorySeq [16], an application of the classical Luria-Delbrück fluctuation test [36], to assay the responses of clonal populations of immortalised Bone Marrow Derived Macrophages (iBMDMs). We found that after approximately 15 cell divisions, 7% of TLR-dependent genes (86 out of 1251), including a number of immune chemokine and cytokine genes, exhibited heritable expression patterns. In contrast, only 2%

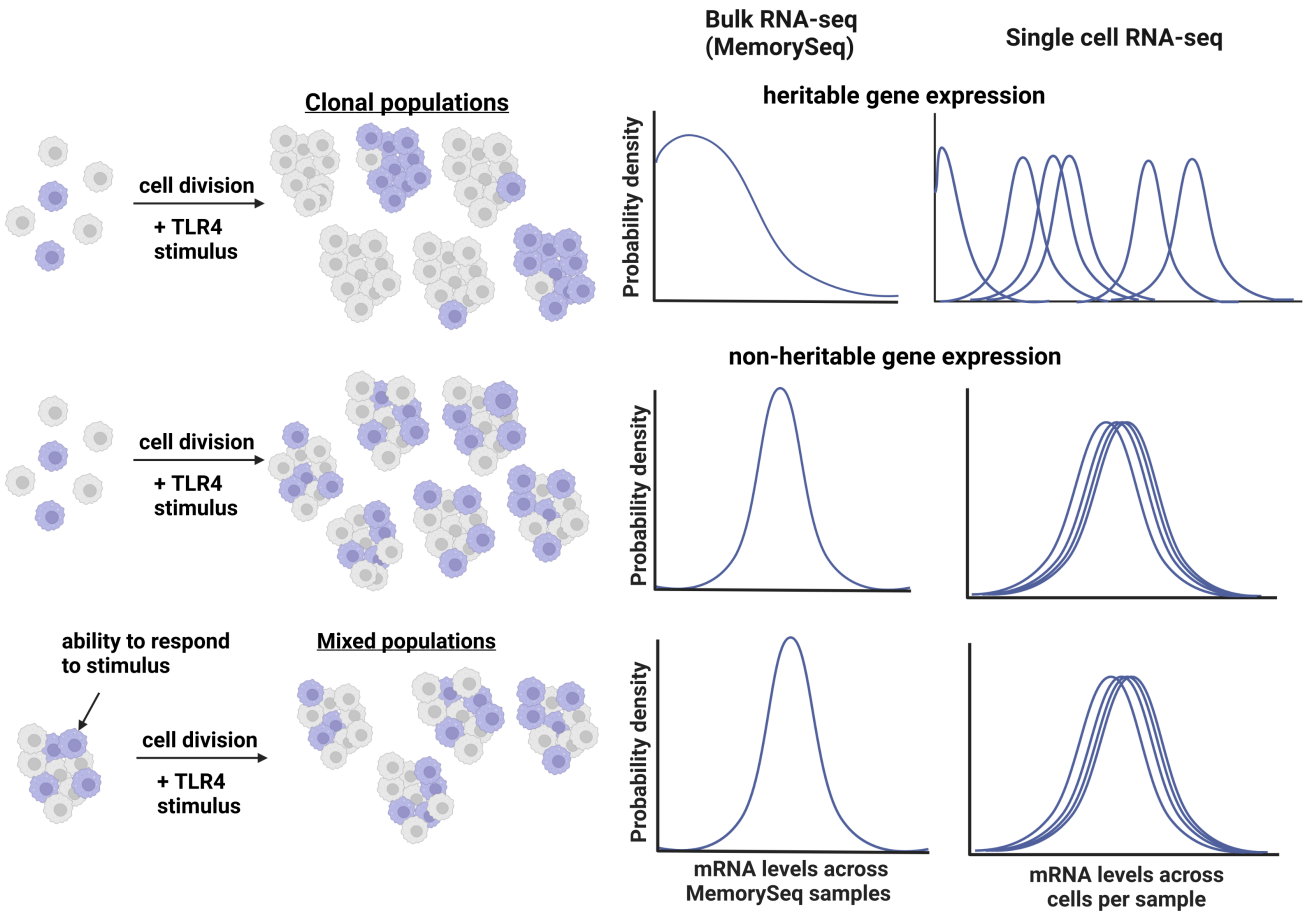


Figure 1: Schematic representation of heritable and non-heritable gene expression in clonal and mixed populations assayed via MemorySeq and scRNA-seq. (**Top**) Heritable gene expression: Stimuli-induced transcription of a hypothetical gene (in purple) is maintained in clonal populations, while some additional fluctuations may occur through consecutive divisions. Populations exhibit wide distributions of MemorySeq bulk samples and large inter-clone variability in the underlying scRNA-seq mRNA distributions. (**Middle**) Non-heritable gene-expression: Stimuli-induced transcription of a hypothetical gene (in purple) fluctuates randomly in clonal populations. Populations exhibit narrow MemorySeq and overlapping scRNA-seq distributions. (**Bottom**) Mixed cell controls exhibit narrow distributions related to sampling/technical noise.

of TLR-independent genes (228 out of 10367) exhibited heritability. Single-cell RNA sequencing (scRNA-seq) of 8 clonal and 4 parental populations demonstrated that individual heritable genes maintained linear mean and variance relationships across clonal populations, which can be attributed to reciprocal transcriptional bursting modulation. Overall, our study demonstrates that single-cell innate immune responses are heritable and provides a theoretical basis for transcriptional control of heritable genes.

Results

MemorySeq reveals heritable TLR-dependent and -independent gene expression patterns

To understand transcriptional heritability in the TLR signalling system we used a combination of genomic approaches relying on the fluctuation test of clonal isogenic populations [36]. If a cell's ability to produce a specific gene expression response (e.g., a rare high expressing cell) is heritable and persist over multiple cell divisions, populations will exhibit large inter-clone variability, such that some populations express a gene with high penetrance (i.e., in majority of cells), while others do not. However, if a gene's expression fluctuates rapidly, responses of clonal populations become identical (subject to measurement error) after few cell divisions, equivalent to that of parental cells (Figure 1). The inter-clonal variability (in comparison to parental cells) has been used as a measure of heritability in previous bulk RNA-seq (MemorySeq) analyses [16], however incorporating scRNA-seq approaches may provide more insights by measuring the whole mRNA distributions in each clone.

To measure heritability in TLR-induced gene expression, we used Fluorescent-Activated Cell Sorting (FACS) to generate MemorySeq clones from isogenic iBMDMs; each clonal population was generated from a single cell grown to approximately 100,000 cells over 15-16 days. To control for technical and sampling noise we sequenced a matching number of parental iBMDM cells, which were seeded into 100,000 populations (referred herein as "mixed populations") and subjected to the same experimental protocol. Prior to bulk RNA sequencing, clonal and mixed populations were

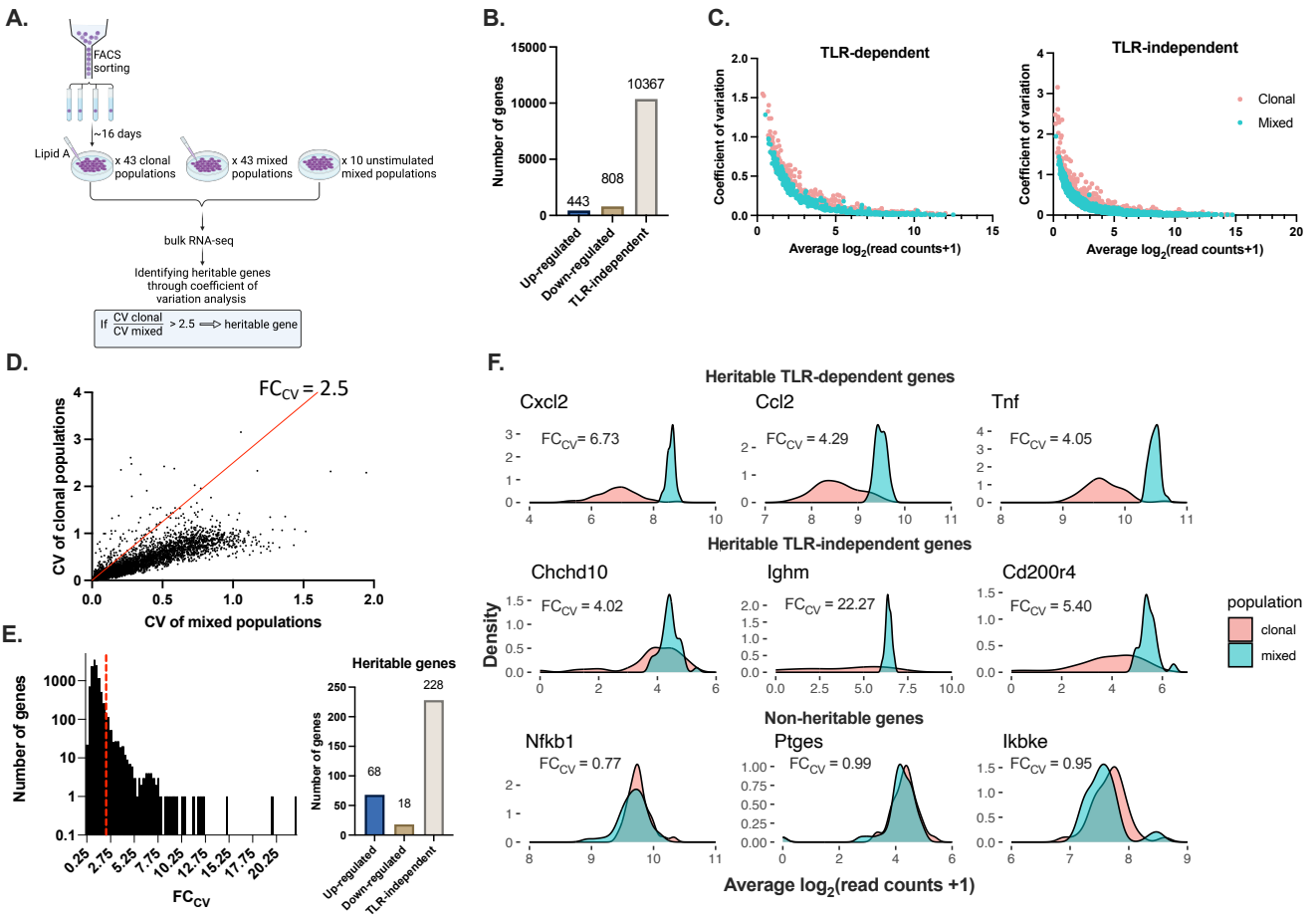


Figure 2: MerySeq reveals heritable gene expression in TLR system A. Schematic representation of MerySeq experiment. Individual iBMDMs cells were FACS-sorted and seeded into in separate wells to grow for a ~16 days (reaching ~100,000 cells). 43 MerySeq clones as well as matching 43 mixed population controls were stimulated with Lipid A for 3 h before RNA sequencing. In addition, experiment involved 10 untreated mixed control samples. Coefficient of variation across clonal and mixed samples was used to identify heritable genes. **B.** The number of genes identified as TLR-dependent (up- and down-regulated) and TLR-independent using mixed populations. **C.** CV versus mean expression levels for the TLR-dependent (left) and TLR-independent genes (right). Each dot represents average expression value for a gene across its clonal (in pink) and mixed samples (in turquoise). **D.** CV of mixed versus clonal samples for the 11618 genes from MerySeq experiment. Red line corresponds to $CV_{clonal} = 2.5\ CV_{mixed}$. **E.** Number of heritable genes. (Left) histogram of the FC_{CV} (fold change CV of clonal samples vs. CV of mixed samples) for the 11618 genes. Broken red line corresponds to the heritability fold change threshold = 2.5, 314 genes passed this threshold. (Right) number of heritable genes identified as TLR-dependent (up- and down-regulated) and TLR-independent. **F.** Examples of heritable (top and middle rows) and non-heritable (bottom) genes. Shown are distributions of expression levels for clonal (in pink) and mixed samples (in turquoise) from MerySeq. The value of FC_{CV} between clonal and mixed is stated under the gene's name.

stimulated for 3h with 500 ng/ml of lipid A, the dominant cytotoxic component of lipopolysaccharides of Gram negative bacteria [37].

We started with 96 samples (45 stimulated clonal populations, 45 stimulated and 10 unstimulated mixed populations), but following library preparation, sequencing, and quality control, we obtained 41 stimulated MerySeq clones, as well as 43 stimulated and 10 untreated mixed populations (Figure 2A). The principal component analyses (PCA) demonstrated an excellent separation between conditions (except of one clonal sample which was subsequently removed, Figure supplement 1) with high coverage of at least 500,000 reads per sample (average of 2,9 million reads). Differential expression analysis of mixed cell populations showed that expression of 1251 genes were significantly affected by lipid A (of which 443 up-regulated and 808 down-regulated, referred to as TLR-dependent), while the expression of 10367 genes were not affected by the treatment (referred to as TLR-independent genes set, Figure 2B).

We used coefficient of variation (CV, i.e., standard deviation normalized by the mean of read counts) to measure variability of expression across individual genes. We found that higher gene expression was associated with reduced variability, while a subset of genes exhibited larger inter-clone variability than that of the technical noise (i.e., CV calculated across mixed populations, Figure 2C). We defined transcriptional heritability using a threshold of 2.5-fold change in CV (FC_{cv}) between MerySeq clones and mixed population controls (consistent with a previous study [16]) (Figure 2D) and found 314 heritable genes in the dataset (Figure 2E). Of the genes identified as heritable, 86 were TLR-dependent (including 68 of upregulated genes, 7% and 15%, respectively), while 228 heritable genes (2%) were TRL-independent (see Table S1 for respective gene lists and Figure 2F for example genes).

Genes identified as heritable in our assay did not gener-

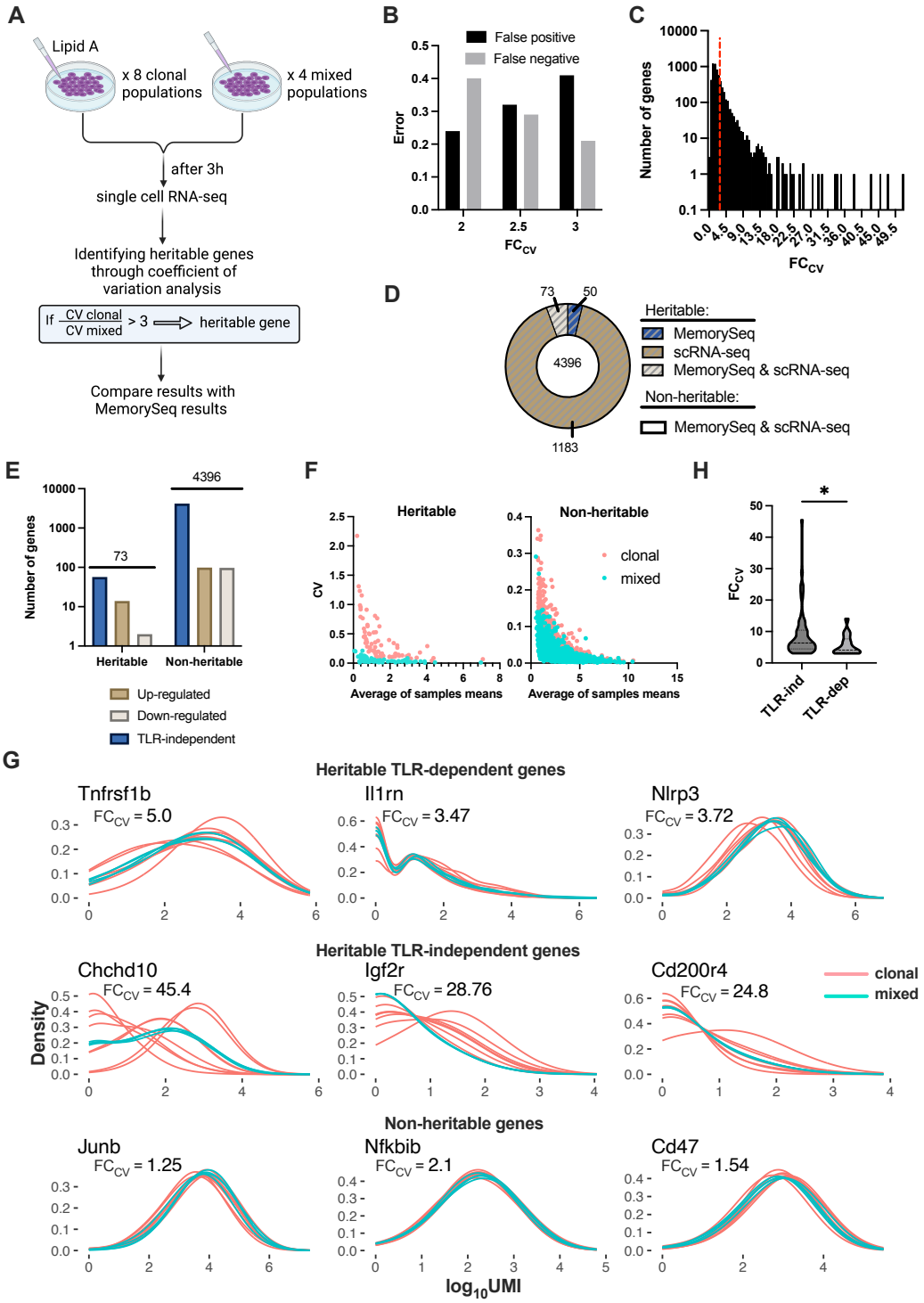


Figure 3: scRNA-seq analysis of heritability. A. Schematic representation of scRNA-seq experiment, involving 8 MemorySeq clones and 4 mixed populations stimulated with lipid A for 3 h. Analysis of CV across clonal samples versus stimulated mixed samples was carried out to identify heritable immune genes using 3-fold change threshold, which were subsequently compared to MemorySeq data. **B.** Sensitivity analysis of heritability index FC_{CV}. Shown is the analysis of false-positive and false-negative errors in identification of heritable genes in scRNA-seq data, given MemorySeq inference, as a function of different FC_{CV} levels (2, 2.5 and 3). **C.** Histogram of the heritability index (FC_{CV}) of clonal vs. mixed samples for the 5702 genes in both scRNA-seq and MemorySeq datasets. Broken red line corresponds to $FC_{CV} = 3$, yielding 1256 genes identified as heritable. **D.** Pie chart showing the number of genes identified as heritable in both experiments (grey), MemorySeq but not scRNA-seq (in blue), scRNA-seq but not MemorySeq (brown), as well as non-heritable in both experiments (in white). **E.** Number TLR-dependent (up- and down-regulated) and -independent genes simultaneously identified as heritable and non-heritable in both MemorySeq and scRNA-seq experiments. **F.** Coefficient of variation versus average of mean expression levels across samples for the 73 genes identified as heritable in both MemorySeq and scRNA-seq experiments (left) and the 4396 non-heritable genes (right). Each dot represents average expression value for a gene across the means of its clonal samples (in pink) and mixed samples (in turquoise). **G.** Examples of heritable (top and middle rows) and non-heritable (last row) genes from E. Shown are scRNA-seq gene expression distributions for clonal (in pink) and mixed (in turquoise) populations. FC_{CV} highlighted next to the gene's name. **H.** Comparison between FC_{CV} of TLR-dependent and -independent 73 heritable genes. Statistical significance assessed with Mann-Whitney test (* p-value<0.05).

ally match previously identified resistance-associated heritable genes in non-phagocytic cell lines [16, 18], but were related to innate immune processes represented by macrophages. Heritable TLR-dependent genes included inflammatory chemokines (Cxcl2, *Ccl2, Ccl3, Ccl7*) and cytokines (*Tnf* α, *Il6*, *Il1*β, *Csf1* and *Csf2*) (Figure 2F), as well as their cognate receptors (*Cclr2* and *Tnfr2*). *Nlrp3*, the critical component of the inflammasome [38], as well as key inflammatory mediator COX-2 encoded by *PTGS2* gene [39] were also identified as heritable.

In terms of the TLR signalling circuitry [40], the key NF-κB family members (*Rel, RelB, NFKB1, NFKB2*), genes encoding their inhibitory κB proteins (*NFKBIA, NFKBIB and NFKBIE*), cognate receptors (including *TLR1, 2, 3, 6, 7, 9*) as well as downstream inhibitory *κβ* kinases (*IKBK, IKBKG*) did not pass heritability threshold. However, we found evidence for heritability of the inducible *TNFAIP3* gene encoding a dual-function deubiquitinase and E3 ligase enzyme A20, critical in controlling multiple key aspects of NF-κB signalling responses [17, 41–43].

No heritability was identified for members of the Janus kinase (JAK)-Signal Transducer and Activator of Transcription (STAT) system [44], including *STAT1, 2, 3, 5* and *6*, *JAK1, 2* and *3*, and type I and II interferon receptors (expression of which was independent of TLR stimulation). In contrast, the inducible suppressor of cytokine signalling 3 (*SOCS3*) gene, a known feedback-inhibitor of JAK-STAT signalling [40], was identified as heritable.

In terms of TLR-induced interferon signalling, previously associated with heterogenous innate immune responses [5, 14], the production of interferon (*IFNB1*) was not heritable and the expression of interferon regulatory factors (*IRF3, 7, 8, 9*) was not heritable, with a notable exception of IRF1, a gene that plays multiple roles in the controlling of innate immune responses [45]. The heritable TLR-independent genes found to be involved in extracellular matrix interactions (*ITGA5, 6, 7, 8*) [46], the antibacterial complement system (*C1ac, C1qb, C1qa, C1ra*), as well as many immune mediators, including those involved in macrophage differentiation and signalling (see Table S2 for ontology analysis).

scRNA-seq demonstrates distinct inter-clone distributions of heritable genes

Bulk analyses of MemorySeq clones identified over 300 genes that passed the heritability threshold, thus suggesting that single-cell TLR-responses are heritable. To validate these analyses, we used scRNA-seq approaches to directly measure mRNA distributions across individual clonal cells and compare against bulk-cell MemorySeq data. We employed a multiplexing protocol (see Methods) to simultaneously assay responses of 8 rederived MemorySeq clones and 4 mixed populations; both stimulated with lipid A for 3 h (Figure 3A). Following quality control, we obtained a total of 9166 single cells, on average 764 (± standard deviation of 166) per population with a median of approximately 56,000 unique molecular identifiers (UMIs) per cell on average across different populations (and median of ~6,900 genes per cell and total ~20,000 genes detected). t-distributed stochastic neighbour embedding (t-SNE) showed that individual clonal and mixed cells clustered together (Figure supplement 2). We found that these cells can be broadly stratified by the cell cycle stage, with 69% of cells corresponding to the longest G1 (Gap 1) stage (Figure supplement 3A and B). We also found that 10% of cells were in the synthesis (S) stage, while remining cells were in the G2M stage (defined as a combined Gap 2 and mitotic phase), consistently with previously measured cell cycle distributions [47]. Notably, the cell cycle distribution was homogenous across the different scNRA-seq populations (Figure supplement 3C).

Based on the scRNA-seq distributions we calculated expression averages across the different populations and identified 5794 genes that were robustly expressed in the dataset, of which 5702 genes that were also expressed in the MemorySeq dataset. Consistently with previous analyses, we then calculated heritability index FC_{cv} (the ratio of the CV between the clonal and mixed populations, respectively) for all 5794 (Table S3). Distribution of FC_{cv} demonstrated many genes with substantial inter-clone variability, thus potentially heritable, in the subset of 5702 genes (Table S4).

While the scRNA-seq data captured entire mRNA distributions across samples (in contrast to sample averages provided by bulk-cell MemorySeq), statistical inference can be generally affected by the limited sample size (i.e., 8 and 4 scRNA-seq vs. >40 MemorySeq samples, respectively). To mitigate against potential sampling errors, we tested different heritability thresholds in relation to reproducibility between the datasets. We found that 3-FC_{cv} threshold in the scRNA-seq analysis reduced the number of false-negative genes (i.e., genes found heritable in more robust Memory-seq, but not in scRNA-seq) to approximately 20%, while allowing the number of false-positives (i.e., genes not heritable in Memory-seq, but identified as heritable in scRNA-seq) to increase (Figure 3B). This yielded 1256 genes in the scRNA-seq dataset that passed the heritability threshold (Figure 3C), with 73 genes being identified as heritable in both datasets (Figure 3D). This included 16 genes, which were TLR-dependent (including 14 upregulated by lipid A), as well as 57 genes that were TLR-independent (Figure 3E). Overall, there was a 78% (total of 5702 genes) agreement between the two datasets in terms of the number of heritable (73) and non-heritable (4396) genes. Individual genes' CVs demonstrated much higher inter-clone variability for the heritable genes, in comparison to mixed controls, and substantially higher to that of non-heritable genes (Figure 3F). The individual mRNA distributions were also consistent with the analyses based on heritability index; mixed cell distributions showed a large degree of overlap between the samples for all genes, with great similarities to clonal distributions of non-heritable genes, while clonal distributions of heritable genes showed variations in the shape or mode (Figure 3G).

Among the heritable TLR-dependent genes identified in both datasets were *Tnfrsf1b, Il1rn, Csf1* as well as *Nlrp3* involved in acute immune responses, while TLR-independent genes included CD36, among others (Figure 3G, see also Table S5 for list of all heritable genes). Interestingly, we found that there was a significant difference in the heri-

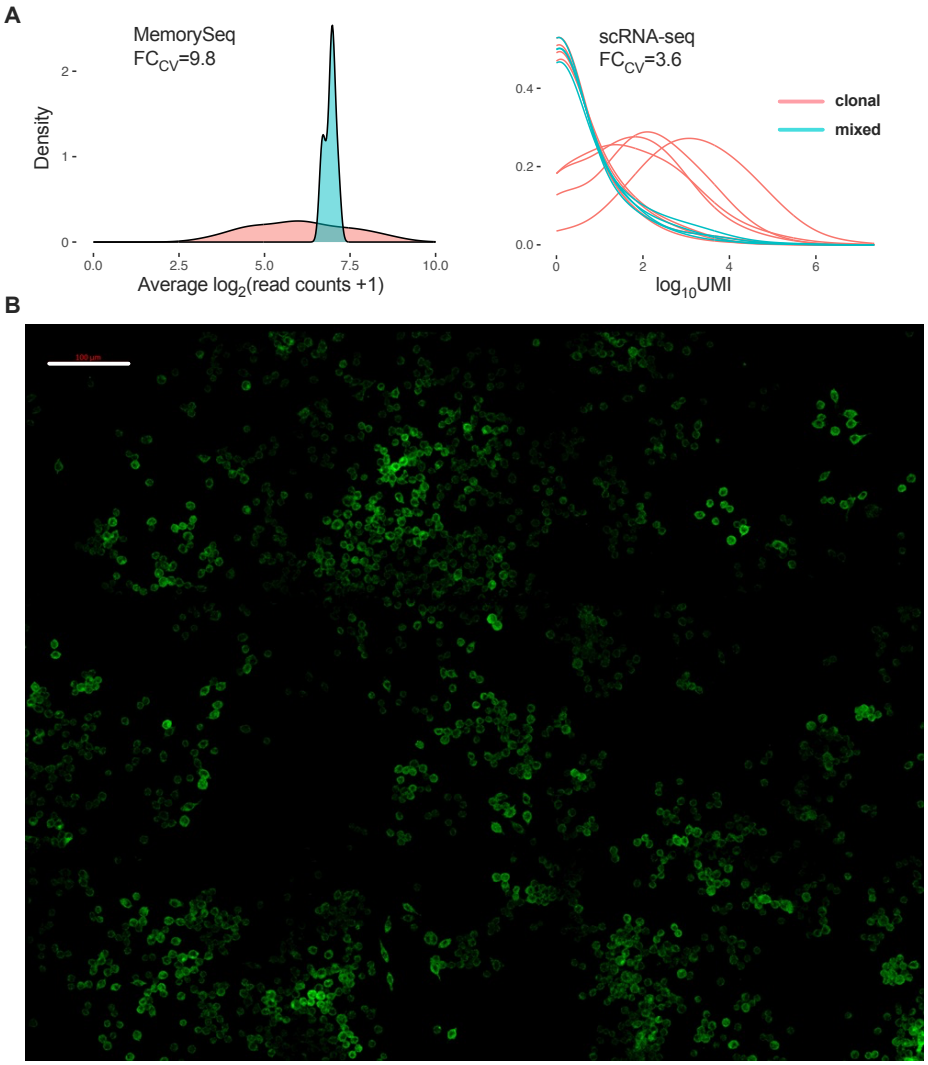


Figure 4: Heritability analysis of Cd36. A. Distributions of Cd36 expression for clonal (in pink) and mixed samples (in turquoise) from MemorySeq (left) and scRNA-seq (right). Displayed on the graph is the value of FC_{CV} between clonal and mixed populations. **B.** Representative immunostaining analysis of Cd36 protein expression in iBMDMs. Cells plated and grown on imaging dish for 2 days before staining. Scale bar 100 µm. Image representative of two replicates.

tability index between the TLR-dependent and independent genes (Figure 3H), suggesting that TLR-dependent genes on average exhibit less memory (which is highlighted by reduced variability between individual mRNA distributions, Figure 3G).

TLR-independent Cd36, encoding platelet glycoprotein 4, is involved in bacterial recognition and phagocytosis in macrophages [48]. Cd36 was characterised by one of the highest heritability indexes (FC_{cv}=9.8) in MemorySeq data and scRNA-seq (FC_{cv}=3.6). The individual mRNA distributions showed large inter-clone variability (Figure 4A). A subset of four clonal distributions were almost identical with that of the mixed populations, characterised by long tails, indicative of the presence of a small subset of high expressing cells in the population. The remaining four distributions showed more symmetrical probability density functions, which centred around higher expression values, suggesting that majority of cells (or at least a much higher proportion than that of other populations) express Cd36 in those populations. These were likely derived from low proportion of high expressing cells present in parental populations. To better validate these analyses, we used immunostaining to measure Cd36 protein expression in population of iBMDMs. Cells were seeded in low density and grown for 2 days, such that daughter and granddaughter cells were spatially colocalised on the imaging dish. Immunostaining demonstrated that Cd36 exhibited a range of protein expression levels across population, with relatively rare high expressing cells (Figure 4B). Importantly, we found that these high expressing cells tend to co-localise together, consistent with transcriptional heritability (Figure 4B).

Overall, our scRNA-seq and immunostaining analyses confirm heritability of gene expression profiles observed by MemorySeq approaches.

Clonal populations maintain expression variability of heritable genes

The fundamental trait of transcriptional heritability is that mRNA distributions exhibit increased inter-clone variability compared to that of the mixed cell populations. While the latter corresponds to the technical noise associated with the scRNA-seq assay, the differences between mRNA distributions across clonal populations may fundamentally capture the differences between clone-specific regulatory events. We therefore investigated if patterns of heritable gene expression share common response characteristics across different populations. We previously demonstrated that TLR-response genes exhibit linear relationships between mean and variance of the mRNA distributions across a range of ligands, doses, times, and cell systems [4, 31]. This relationship demonstrates that while cells adjust their mean mRNA response to changes of stimulus, they simultaneously maintain the overall cell-to-cell variability level. Consequently, the heterogeneity of a particular gene response, i.e., variance normalised by the mean expression, is constant and defined by the slope of the regression line. Given this notion of transcriptional variability control, we therefore asked if transcriptional heterogeneity of heritable genes exhibit similar behaviour.

First, we analysed the dispersion of the mRNA mean and variance across clonal and mixed cell populations. We found that the 73 heritable genes exhibited significantly higher dispersion compared to that of non-heritable genes (measured as a fold change of the centroid distance between clonal and mixed populations, Figure 5A). In the case of non-heritable genes or mixed cell populations, the dispersion was generally small, consistent with overlapping mRNA distributions, subject to technical noise. To uncover patterns in the dispersion across clonal populations of heritable genes, we applied regression analyses to test if the variance of mRNA distributions is a linear function of the mean response. We found that 1543 out of the 4396 non-heritable genes (35%) could be fitted with a linear mean-variance relationship when pooling together all mixed and clonal samples per gene (as determined by significant regression slope, p-val<0.05). Only a subset of 465 genes (11%) exhibited a high confidence fit (coefficient of determination, $R^2 > 0.6$, Figure 5B). On the other hand, 50 out of the 73 heritable genes (68%) were fitted with a linear regression (p-val<0.05) with 42 (57%) passing high confidence threshold ($R^2 > 0.6$, Figure 5B, also see Table S6 for list of genes).

We subsequently asked whether two different linear relationships, one for clonal populations and another for mixed populations, can be fitted, testing whether different populations exhibit different noise control. However, we could only statistically distinguish fitted regression slopes for 3 (5% of total) heritable and 51 (<1%) non-heritable genes across clonal and mixed populations (Figure 5C). Given the inherent noise and limited sample size of the data, these analyses suggest that lack of heritability could be interpreted by clonal populations having similar characteristics in terms of the moments of the mRNA distribution to that of the parental cells (see Figure 5D for example relationships for *Nfkbia, Nfkbib, Cd47* and *Junb* genes). In contrast, while heritable genes exhibit different mRNA distributions across different populations, they maintain the overall response variability (as defined by the constant mean-variance relationship for all mixed and clonal samples, see Figure 5E for example genes). We interpret this behaviour by suggesting that transcriptional heritability corresponds to positioning clonal populations onto the gene-specific mean-variance line away from the point corresponding to the moments of the parental population. In this context, as populations loose transcriptional heritability, they collapse to parental populations on the linear relationship (e.g., *Il1rn* in general exhibits less dispersion than highly heritable *Cd36*, Figure 5E).

Finally, we wanted to understand differences in transcriptional regulation of heritable genes across different cell populations. In general, transcriptional heterogeneity has been predominantly attributed to the process of transcriptional bursting, i.e., random changes of gene activity permissive for transcription [49]. In this process, the burst size (i.e., how many mRNAs per 'on' period) and frequency (i.e. how often in 'on' state) control the mRNA output of individual cells, and thus the heterogeneity of the cell population [50]. The moments of the mRNA distributions can be used to determine burst size and frequency, such that burst size $b_s = \sigma^2/\mu$ (i.e., the Fano factor) and burst frequency $b_f = \mu/b_s$, to quantify the departure from 'non-bursty' Poissonian mRNA production, characterised by $b_s = 1$ and $b_f = \infty$ [22, 23, 30, 51]. We have previously shown that linear-mean variance relationships constrain transcriptional bursting by reciprocal modulation of burst size and burst frequency as the function of the mean mRNA response. In particular, given that $\sigma^2 = \alpha\mu + \alpha_0$, bursting characteristics can be derived such that burst size $b_s = \alpha_0/\mu + \alpha$ and burst frequency $b_f = \alpha_0/(b_s(b_s - \alpha))$ [4, 31].

We applied this theory on our heritable genes and found that the heritable gene expression patterns show a similar phenomenon, that is, the location on the mean-variance line corresponds to different, but theoretically predictable, transcriptional bursting characteristics (see Figure 5F for the predicted and observed modulation of transcriptional bursting across mean-variance lines for example genes). In this case, parental populations exhibit similar bursting characteristics to each other, while individual clones exhibit different transcriptional bursting characteristics resulting in unique gene expression patterns. Interestingly, we observed that differences between individual clones were predominantly associated with changes of burst frequency rather than burst size (Figure 5G), suggesting that the former is the underlying control parameter.

The phenomenon of transcriptional bursting involves stochastic regulation, which implies independent regulation of individual genes. In practice, groups of genes may exhibit co-fluctuations, indicative of shared transcriptional regulation such as upstream transcription factors or common control through general regulatory networks [16]. Having identified a set of TLR-dependent and independent heritable genes, we therefore tested if the expression of heritable genes co-fluctuates. If a particular clonal population

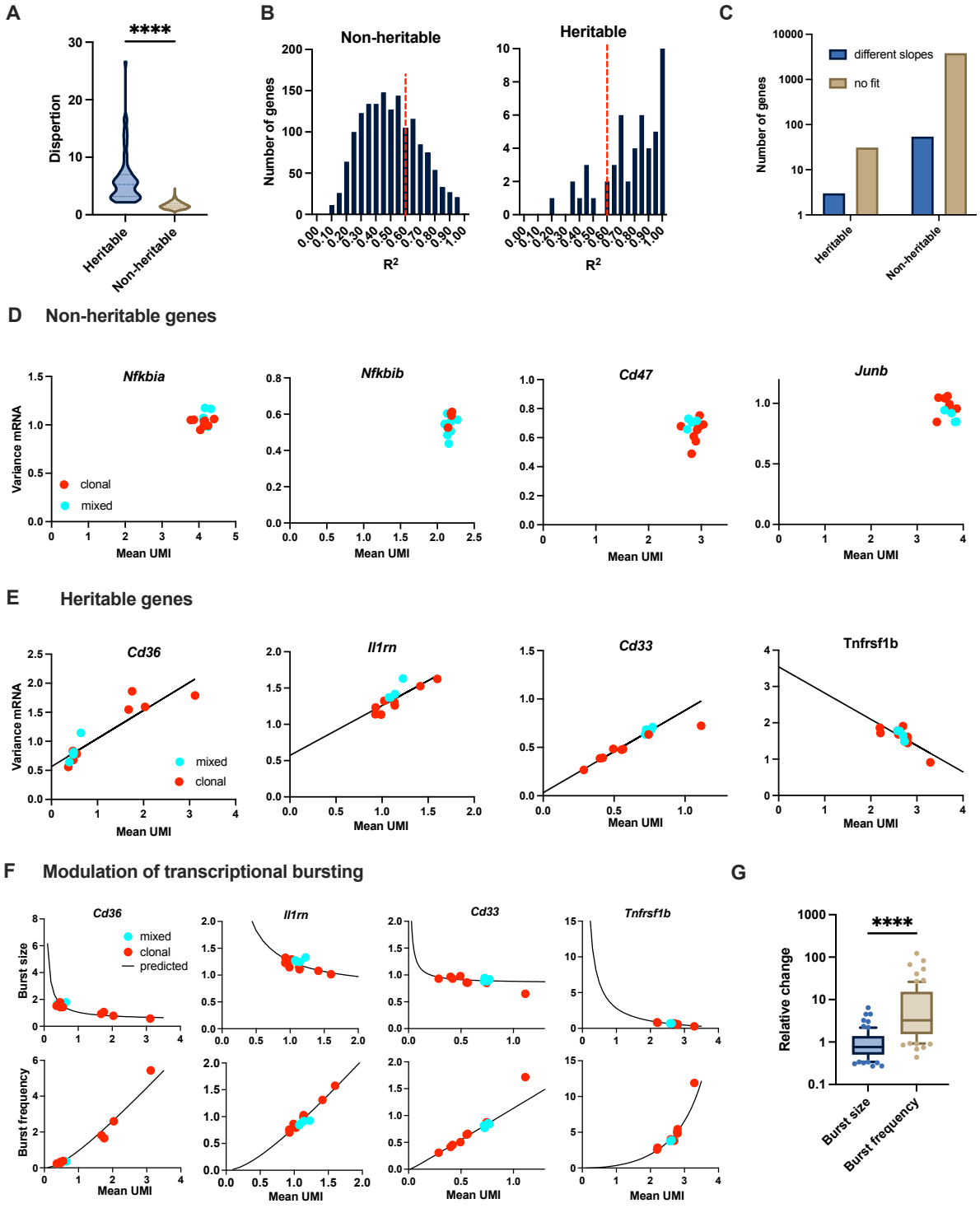


Figure 5: Heritable genes maintain response heterogeneity A. Dispersion in the moments of the mRNA distribution between heritable and non-heritable genes. Shown is the comparison of fold change in centroid distance of clonal samples to that of mixed samples for heritable and non-heritable genes. Statistical significance between groups is assessed with Mann-Whitney test (**** p-value <0.0001). **B.** Histogram of coefficient of determination (R^2) for the 1543 non-heritable (left) and 50 heritable (right) mean-variance linear fits characterised by a significant regression slope (p-value < 0.05). $R^2 = 0.6$ (broken line in red) corresponds to the high confidence gene cut-off. 42 heritable genes (57%) passed this cut-off compared to 465 non-heritable genes (11%). **C.** Number of genes that fit mean-variance linear relationship with different slopes for clonal and mixed (blue) and those that do not fit any mean-variance relationship (no fit) for heritable vs. non-heritable genes. **D.** Examples of mean-variance relationship for non-heritable genes. Individual cell populations (clonal and mixed) characterised with differentially coloured points (as per legend). **E.** Examples of linear mean-variance relationship for heritable genes. Individual populations characterised with differentially coloured points (as per legend). Solid line corresponds to fitted mean-variance regression line across all populations. **F.** Burst size and frequency modulation along the linear relationships for genes in E. Moment estimators of burst size and frequency for individual populations represented with differentially coloured points (as per legend). Solid curve corresponds to predicted modulation calculated based on the fitted mean-variance relationship. **G.** Relative change of burst size and burst frequency across 73 heritable genes. Shown are individual changes defied as (max-min)/min across genes with population mean and standard deviation. Statistical significance between groups is assessed with paired Wilcoxon test (**** p-value <0.0001).

or single cell has a high abundance of a specific heritable gene, then the abundance of co-fluctuating transcript will be correlated. We therefore calculated pairwise Pearson's correlation coefficients between the expression of all 73 heritable genes across MemorySeq and scRNA data. We found evidence for blocks of co-regulated genes, which exhibited high correlations, both in MemorySeq as well as in more noisy scRNA-seq (Figure supplement 4A). In general, all TLR-dependent genes, except for *Napepld* exhibited high cross-correlations, likely reflecting robust stimuli-induced activation of these genes (Figure supplement 4B). However, we also detected groups of co-fluctuating TLR-independent genes (Figure supplement 4C), which included genes involved in the complement system (*C1qc* and *C1qb*) as well as a number of transmembrane proteins (*Cd36, Cd33, Cd200r4*) among others. Overall, these analyses demonstrate an additional layer of regulation enabling co-regulation of heritable genes.

Discussion

The activation of the innate immunity is an inherently heterogenous process, with genetically identical cells exhibiting different ability to produce key inflammatory mediators [11, 47] and ultimately to control the pathogen [14, 52, 53]. Previous work suggested that immune signalling involves stochastic regulation [4, 8, 29, 30, 54, 55], however some seemingly stochastic responses can be explained by epigenetic control [34], cell cycle state [47], as well as cellular [17, 56] and population context [57]. Recent analyses suggest that gene expression patterns can be generally inherited through dividing cells across many generations [16] [32]. Here we used genomic approaches to investigate whether single-cell transcriptional responses of the evolutionarily-conserved TLR system exhibit transcriptional heritability. We used MemorySeq [16], an assay relying on comparisons between responses of clonal populations, and demonstrated that 7% of TLR-dependent genes (and 15% of upregulated genes), including a number of immune cytokines and chemokines, exhibited heritable expression patterns.

Transcriptional heritability has been generally described as a transient phenomenon [6, 16, 32] with the clonal characteristics eventually converging to that of parental cells. In this work we focused on long-term heritability, associated with approximately 15-16 cell divisions due to feasibility of the genomic approaches. We found that the heritability index (the ratio of inter-clone variability to that of mixed cell populations) is generally lower in TLR-dependent genes compared to TLR-independent genes, suggesting that immune response in macrophages is generally more transient. For example, TLR-independent genes *Chchd10, Cd36* and *cd200r4* include 'high' and 'low' clones across different populations, while TLR-dependent genes exhibit less distinct patterns in the scRNA-seq data (Figure 3G). These constitutively expressed genes, however, contribute to the overall immune responses by controlling various aspects of cell signalling. For example, heterogenous and heritable expression of Cd36 might be related to individual-cell rate of phagocytosis and pathogen control [48]. Given the transient nature of heritability, it is however likely that many more TLR-regulated genes described in this work exhibit heritable expression, but this heritability disappears after just a few cell divisions. Primary bone marrow-derived and tissue-resident macrophage lineages undergo proliferation inflammatory responses [58], suggesting that patterns described here might contribute substantially to control of innate immune responses in vivo. Similarly, heritable traits might contribute to innate immune memory, i.e., ability of cells to retain memory of prior infections to produce robust responses upon reinfection [59]. One intriguing possibility is that cells that deal well with pathogens and survive an infection are more likely to proliferate and pass on their heritable gene expression to manage subsequent infections.

Our scRNA-seq data surprisingly demonstrates that individual clones maintain variability of the heritable gene expression. We show that while an individual population of clones exhibits a different level of expression to parental (i.e., 'mixed') populations, the overall variance of the mRNA response is linearly constrained by the mean mRNA level. Under this relationship, noise characteristics in clonal populations diverge from that of the parental cells, with the latter forming a single point (subject to technical noise) on the gene's mean-variance line. One interpretation of this phenomenon is that clonality can be interpreted with departure from the "noise equilibrium" represented by parental cells, with individual genes returning to this equilibrium at different timescales. We showed that individual clones differ in their transcriptional bursting characterising, which overall are constrained by the gene-specific mean-variance relationship [4, 31]. In fact, we demonstrate that differences between individual clonal populations can be predominately attributed to the frequency modulation. Previous work suggests that clonality and heritability are associated with long-term epigenetic control [6], including chromatin accessibility [60] and methylation states [61, 62]. While the invariant mean-variance relationship is likely to be a property of the gene specific encoding of burst kinetics by promoters and enhancer elements [27, 28], how heritable expression is mechanistically controlled in individual cells and groups of co-fluctuating genes remains unclear. Frequency modulation was previously associated with changes of histone acetylation [22], and in general distal regulatory elements in DNA which controlled cell-type specific expression [27]. Recently, single cell TLR-dependent expression was linked to frequency modulation via cohesin-mediated promoter-enhancer coupling [15]. While genes involved in the cohesin complex formation (*SMC1, SMC3, SCC1, SCC3*) [63] are not heritable in our data, this and other epigenetic mechanisms provide avenues for future studies. While current models of transcriptional bursting assume stochastic activation of transcription in homogenous immune cell populations [4, 29–31], existence of heritable gene expression patterns require new modelling approaches, which will need to also account for transitions between transient heritable states.

Overall, this work demonstrates that the acute innate immune gene expression responses of the TLR system are heritable over multiple generations and provide a theoretical basis for this pthenomenon.

Materials and Methods

Cell lines and culture

Immortalised bone marrow derived macrophages (iBMDMs) [64] were cultured in Dulbecco's modified Eagle's medium (DMEM) high glucose with sodium pyruvate and sodium bicarbonate (Sigma-Aldrich D6429) supplemented with 10% heat inactivated foetal bovine serum (FBS; Gibco 10500064). All incubations were done at 37°C 5% CO2 and all cell handling under laminar air flow unless otherwise stated.

Preparation of clonal populations

FACS was used to sort single iBMDMs into prewarmed DMEM+FBS in 96 well plates and incubated to generate clonal populations. To reduce clumping during growth from single cells, cells were dispersed by gently reverse pipetting on days 5 and 8. When cells were 60-80% confluent (day 12-14) clonal populations were transferred to 6 well plates and grown until reached 100,000-200,000 cells per plate (day 15-16) for subsequent experiments.

Lipid A stimulation

Lipid A diphosphoryl from Salmonella enterica serotype Minnesota Re595 (Sigma-Aldrich L0774) stock was prepared at 500mg ml-1 in 50% DMSO and stored at -20°C until use. Clonal or mixed populations in 6 well plates at approximately 50% confluence were stimulated with lipid A at a final concentration of 500 ng ml^{-1} for 3h. Unstimulated controls were treated with an equivalent volume of 50% DMSO for 3h.

MemorySeq library construction, sequencing, and differential expression analysis

43 stimulated clonal populations, 43 stimulated mixed populations and 10 unstimulated mixed populations were prepared. Cells were harvested by scraping and stored in RNA*later* (Invitrogen AM7020) for stable storage. Cells were pelleted by centrifugation and RNA*later* removed prior to RNA extraction with the RNeasy Plus mini kit (Qiagen 74136) according to the manufacturer's instructions. RNA was stored at -80°C. Total RNA was submitted to the Genomic Technologies Core Facility (GTCF). Quality and integrity of the RNA samples were assessed using a 4200 TapeStation (Agilent Technologies) and then libraries generated using the Illumina® Stranded mRNA Prep. Ligation kit (Illumina, Inc.) according to the manufacturer's protocol. Briefly, total RNA (typically 0.025-1ug) was used as input material from which polyadenylated mRNA was purified using poly-T, oligo-attached, magnetic beads. Next, the mRNA was fragmented under elevated temperature and then reverse transcribed into first strand cDNA using random hexamer primers and in the presence of Actinomycin D (thus improving strand specificity whilst mitigating spurious DNA-dependent synthesis). Following removal of the template RNA, second strand cDNA was then synthesized to yield blunt-ended, double-stranded cDNA fragments. Strand specificity was maintained by the incorporation of deoxyuridine triphosphate (dUTP) in place of dTTP to quench the second strand during subsequent amplification. Following a single adenine (A) base addition, adapters with a corresponding, complementary thymine (T) overhang were ligated to the cDNA fragments. Pre-index anchors were then ligated to the ends of the double-stranded cDNA fragments to prepare them for dual indexing. A subsequent PCR amplification step was then used to add the index adapter sequences to create the final cDNA library. The adapter indices enabled the multiplexing of the libraries, which were pooled prior to clustering on a flow-cell from a High-Output NextSeq 500/550 v2.5 kit. The loaded flow-cell was then paired-end sequenced (76 + 76 cycles, plus indices) on an Illumina NextSeq500 instrument. Finally, the output data was demultiplexed and BCL-to-Fastq conversion performed using Illumina's bcl2fastq software, version 2.20.0.422.

Unmapped paired-reads of 74bp from an Illumina NextSeq500 sequencer were interrogated using a quality control pipeline consisting of FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FastQ Screen v0.13.0 (https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/). The reads were trimmed to remove any adapter or poor quality sequence using Trimmomatic v0.39 [65]; reads were truncated at a sliding 4bp window, starting 5', with a mean quality <Q20, and removed if the final length was less than 36bp. Additional flags included: 'ILLUMINACLIP:./Truseq3-PE-2_Nextera-PE.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:36'.The filtered reads were mapped to the mouse reference sequence (mm10/GRCm38) from the UCSC browser [66], using STAR v2.7.7a [67]. The genome index was created using the comprehensive mouse Gencode vM25 gene annotation [68] applying a read overhang (–sjdbOverhang 75). During mapping the flags '–quantMode GeneCounts' was used to generate read counts into genes.Normalisation and differential expression analysis was performed using DESeq2 v1.30.1 [69] on R v4.0.4. Log fold change shrinkage was applied using the lfcShrink function along with the "apeglm" algorithm [70].

Following mapping and analyses we retained all but 2 clonal samples with a depth of at least 500,000 reads per library (2.9 million reads per sample on average). One sample was removed due to poor coverage, while another (sample 18) was subsequently removed because it clustered with mixed cell populations, see Figure supplement 1 for the principal component analysis (PCA).

scRNA-seq single cell isolation, library construction and sequencing

8 stimulated clonal populations and 4 stimulated mixed populations were prepared. Cells were harvested by scraping

then clonal populations were labelled prior to pooling according with the 10x genomics 3' CellPlex kit according to the manufacturer's instructions for >80% viable cells (protocol 1 document number CG000391). Library preparation and sequencing was performed by the University of Manchester Genomic Technologies Core Facility. Gene expression and Cell Multiplexing libraries were prepared from Cell Multiplexing Oligo labelled cells using the Chromium Controller and Single Cell 3' Reagent Kits v3.1 (10x Genomics, Inc. Pleasanton, USA) according to the manufacturer's protocol (CG000388 Rev B). Briefly, nanoliter-scale Gel Beads-in-emulsion (GEMs) were generated by combining barcoded Gel Beads, a master mix containing cells, and partitioning oil onto a Chromium chip. Cells were delivered at a limiting dilution, such that the majority (90-99%) of generated GEMs contain no cell, while the remainder largely contain a single cell. The Gel Beads were then dissolved, primers released, and any co-partitioned cells lysed.

Primers containing an Illumina TruSeq Read 1 sequencing primer, a 16-nucleotide 10x Barcode, a 12-nucleotide unique molecular identifier (UMI) and a 30-nucleotide poly-(dT) sequence were then mixed with the cell lysate and a master mix containing reverse transcription (RT) reagents along with primers containing an Illumina Nextera Read 1, a 16-nucleotide 10x Barcode, a 12-nucleotide UMI and a capture sequence. Incubation of the GEMs yielded barcoded cDNA from poly-adenylated mRNA and barcoded DNA from the Cell Multiplexing Oligo Feature barcode. Following incubation, GEMs were broken, and pooled fractions recovered. The cell barcoded cDNA was then purified from the post GEM-RT reaction mixture using silane magnetic beads and amplified via PCR to generate sufficient mass for library constructions. The amplified cDNA molecules for 3' Gene Expression were separated from those for Cell Multiplexing library construction by size selection. For the 3' Gene Expression library enzymatic fragmentation and size selection were then used to optimize the cDNA amplicon size. Illumina P5, P7, i7 i5 sample indexes, and TruSeq Read 2 sequence were added via end repair, A-tailing, adaptor ligation, and PCR to yield final Illumina-compatible sequencing libraries. For the Cell Multiplexing library P5, P7, i7 i5 sample indexes and Nextera Read 2 were added via PCR to amplified DNA from Cell Multiplexing Oligo Feature Barcodes to yield final Illumina-compatible sequencing libraries.

Single Cell 3' libraries comprised standard Illumina paired-end constructs flanked with P5 and P7 sequences. The 16 bp 10x Barcode and 12 bp UMI were encoded in Read 1, while Read 2 was used to sequence the cDNA fragment in 3' Gene Expression libraries while read 2N was used to sequence the DNA from Cell Multiplexing Feature barcode. Sample index sequences were incorporated as the i7 and i5 index read. Paired-end sequencing (28:90) was performed on the Illumina NovaSeq 6000 platform. The .bcl sequence data were processed for QC purposes using bcl2fastq software (v. 2.20.0.422) and the resulting .fastq files assessed using FastQC (v. 0.11.3), FastqScreen (v. 0.14.0) and FastqStrand (v. 1.11.1) prior to pre-processing with the CellRanger pipeline.

scRNA-seq data processing, cell filtering and cell cycle assignment

The 10x Genomics Cell Ranger pipeline (v7.0.0) was used to process raw sequencing data. The base call (BCL) files produced by the sequencer were demultiplexed and converted to FASTQ files using "cellranger mkfastq". The gene expression and multiplexing capture FASTQ files were processed using "cellranger multi" and mapped against the pre-built Mouse reference package from 10X Genomics (mm10-2020-A) to generate the per-sample gene-cell barcode matrix. The single-cell data were processed in R environment (v4.1) following the workflow documented in Orchestrating Single-Cell Analysis with Bioconductor [71]. Briefly, for each sample, the HDF5 file generated by Cell Ranger was imported into R to create a SingleCellExperiment object. A combination of median absolute deviation (MAD), as implemented by the "isOutlier" function in the scuttle R package (v1.4.0) and exact thresholds was used to identify and subsequently remove low quality cells before data integration. Cell cycle phase classification was performed using the "cyclone" function and pre-trained classifiers from the scran R package (v1.22.1) and obtained the predicted phase for each cell.

scRNA-seq data integration, visualisation, and cell clustering

The "multiBatchNorm" function from the batchelor R package (v1.10.0) was used to re-compute the log-normalized expression values of the combined single-cell data. Mutual nearest neighbours (MNN) approach available from the batchelor R package was used to perform batch correction on top 2000 highly variable genes. Then, the first 50 dimensions of the MNN low-dimensional corrected coordinates were used as input to produce the t-stochastic neighbour embedding (t-SNE) projection using the "runTSNE" function from the scatter R package (v1.22.0) respectively.

Heritability analysis

Heritability index was defined as the ratio of coefficient of variations (CV) of the $\log_2(\text{read counts}+1)$ in the clonal and mixed samples, respectively. In the MemorySeq analysis this index was calculated for 11619 genes with a total number of read counts >100 across all samples. In the scRNA-seq, the index was calculated based on average UMIs per sample for 5702 genes that were also expressed in MemorySeq (genes with at least one clonal sample with $\log_{10}(\text{UMIs}) > 1$ were considered expressed, with total of 5794 genes passing the threshold). Cut-offs of 2.5 and 3 for MemorySeq and noisier scRNA-seq data, respectively, were applied to define heritable genes. Ontology analysis was performed in Enrichr [72, 73]. For each pair of 73 heritable genes identified in both MemorySeq and scRNA-seq datasets, we calculated the Pearson correlation coefficient between their expression across all samples (clonal and mixed). Correlation matrices were represented as heatmaps with the same order of genes (as per MemeorySeq).

Regression analysis

The mean-variance relationships ($\sigma^2 = \alpha\mu + \alpha_0$) were fitted using the "lmrob" function from the *robustbase* R package (version 0.95-1) ensuring robust linear regression. A model's fit was considered successful if the slope (α) was statistically significant based on p-value < 0.05, and it provided a good overall fit (coefficient of determination $R^2 > 0.6$). Assuming linear constraints of mRNA mean and variance, theoretical transcriptional bursting characteristics were analytically derived, using moment estimators of mRNA count distributions; burst size $b_s = \alpha_0/\mu + \alpha$, burst frequency $b_f = \mu^2/(\alpha_0 + \alpha\mu)$ and $b_f = \alpha_0/(b_s(b_s - \alpha))$, as previously described [4, 31]. Student's t-test was performed to determine whether the two slopes are significantly different.

Centroid dispersion

The centroid-based distance was calculated for heritable and non-heritable genes as follows: First the centroid point coordinates were computed $(x_c, y_c) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i\right)$, where x and y are mean and variance, respectively and n is the number of points (samples). Second, the distance between each point and the centroid was calculated using the Euclidean distance formula $d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$. The overall dispersion was then measured by computing the average between the points and the centroid $\frac{1}{n}\sum_{i=1}^{n} d_i$. Finally, the fold change ratio between clonal and mixed populations were calculated.

Analysis of Cd36 expression

10^4 iBMDM cells were seeded in the 35 mm glass-bottom imaging dish (Griner Bio One) and incubated over 2 days to spatially segregate dividing cells. Cells were fixed with 5 ml of 4% Paraformaldyhde/PBS and incubated at room temperature for 15 mins, and subsequently permeabilised with 250 µl of 0.1% triton X-100 PBS for a 4 mins. 100 µl of the anti-mouse Cd36 antibody (Bio-Rad, MCA2748, clone MF3) diluted in blocking buffer 1:100 was added and incubated for 20 mins in room temperature. 100 µl of the secondary goat antibody (Biolegend) diluted in blocking buffer 1 in a 100 was added and incubated for 20 mins. Z-stack tile scan images were taken on Zeiss LSM 880 confocal microscope. Median filter with 5x5 tile-scan with 12 Z-slices was applied and maximum intensity projection images were collected.

Statistical Analyses

Statistical analysis was performed using GraphPad Prism 9 software. Data was tested for normality using D'Agostino-Pearson test. Two-sample comparisons were conducted using non-parametric Mann Whitney test.

Data Availability

Generated sequencing data have been deposited in the ArrayExpress database at EMBL-EBI under accession number E-MTAB-11041 (https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-11041) and E-MTAB-13014 (https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-13014). Supplementary tables can be found at (https://github.com/nalachkar/MemorySeq-paper.git).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

NA was supported by Wellcome Trust PhD Studentship. This work was also supported by BBSRC (BB/R007691/1).

Author contributions

NA performed computational analyses presented in the manuscript as well as immunostaining. JM conducted genomics experiments. MM, IR, AS and PP provided supervision and conceptualisation. PP with assistance of NA wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Ian Donaldson, I-Hsuan Lin, Claire Morrisroe and Andy Hayes of the Bioinformatics and Genomic Technologies Core Facilities at the University of Manchester for providing support regarding genomics analyses.

References

[1] C. E. Bryant, M. Symmons, and N. J. Gay, Mol Immunol **63**, 162 (2015).
[2] A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park, and A. Regev, Nature **498**, 236 (2013).
[3] Q. Xue, Y. Lu, M. R. Eisele, E. S. Sulistijo, N. Khan, R. Fan, and K. Miller-Jensen, Science Signaling **8** (2015), ARTN ra59 DOI 10.1126/scisignal.aaa2155.
[4] J. Bagnall, W. Rowe, N. Alachkar, J. Roberts, H. England, C. Clark, M. Platt, D. A. Jackson, M. Muldoon, and P. Paszek, Cell Syst **11**, 300 (2020).
[5] A. K. Shalek, R. Satija, J. Shuga, J. J. Trombetta, D. Gennert, D. Lu, P. Chen, R. S. Gertner, J. T. Gaublomme, N. Yosef, S. Schwartz, B. Fowler, S. Weaver, J. Wang, X. Wang, R. Ding, R. Raychowdhury, N. Friedman, N. Hacohen, H. Park, A. P. May, and A. Regev, Nature **510**, 363 (2014).
[6] L. C. Van Eyndhoven, V. P. G. Verberne, C. V. C. Bouten, A. Singh, and J. Tel, eLife **12**, e83055 (2023).
[7] T. Hagai, X. Chen, R. J. Miragaia, R. Rostom, T. Gomes, N. Kunowska, J. Henriksson, J. E. Park, V. Proserpio, G. Donati, L. Bossini-Castillo, F. A. Vieira Braga, G. Naamati, J. Fletcher, E. Stephenson, P. Vegh, G. Trynka, I. Kondova, M. Dennis, M. Haniffa, A. Nourmohammad, M. Lassig, and S. A. Teichmann, Nature **563**, 197 (2018).
[8] P. Paszek, S. Ryan, L. Ashall, K. Sillitoe, C. V. Harper, D. G. Spiller, D. A. Rand, and M. R. White, Proc Natl Acad Sci U S A **107**, 11644 (2010).
[9] J. J. Muldoon, Y. Chuang, N. Bagheri, and J. N. Leonard, Nat Commun **11**, 878 (2020).

[10] J. Bagnall, C. Boddington, H. England, R. Brignall, P. Downton, Z. Alsoufi, J. Boyd, W. Rowe, A. Bennett, C. Walker, A. Adamson, N. M. X. Patel, R. O'Cualain, L. Schmidt, D. G. Spiller, D. A. Jackson, W. Muller, M. Muldoon, M. R. H. White, and P. Paszek, Sci Signal **11** (2018), 10.1126/scisignal.aaf3998.

[11] S. Tay, J. J. Hughey, T. K. Lee, T. Lipniacki, S. R. Quake, and M. W. Covert, Nature **466**, 267 (2010).

[12] E. W. Martin, A. Pacholewska, H. Patel, H. Dashora, and M. H. Sung, Sci Signal **13** (2020), 10.1126/scisignal.aax7195.

[13] A. Adelaja, B. Taylor, K. M. Sheu, Y. Liu, S. Luecke, and A. Hoffmann, Immunity **54**, 916 (2021).

[14] R. Avraham, N. Haseley, D. Brown, C. Penaranda, H. B. Jijon, J. J. Trombetta, R. Satija, A. K. Shalek, R. J. Xavier, A. Regev, and D. T. Hung, Cell **162**, 1309 (2015).

[15] I. Robles-Rebollo, S. Cuartero, A. Canellas-Socias, S. Wells, M. M. Karimi, E. Mereu, A. G. Chivu, H. Heyn, C. Whilding, D. Dormann, S. Marguerat, I. Rioja, R. K. Prinjha, M. P. H. Stumpf, A. G. Fisher, and M. Merkenschlager, Nature Communications **13**, 4342 (2022).

[16] S. M. Shaffer, B. L. Emert, R. A. Reyes Hueros, C. Cote, G. Harmange, D. L. Schaff, A. E. Sizemore, R. Gupte, E. Torre, A. Singh, D. S. Bassett, and A. Raj, Cell **182**, 947 (2020).

[17] A. Adamson, C. Boddington, P. Downton, W. Rowe, J. Bagnall, C. Lam, A. Maya-Mendoza, L. Schmidt, C. V. Harper, D. G. Spiller, D. A. Rand, D. A. Jackson, M. R. H. White, and P. Paszek, Nat Commun **7** (2016), 10.1038/ncomms12057.

[18] S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, E. Eggan, I. N. Anastopoulos, C. A. Vargas-Garcia, A. Singh, K. L. Nathanson, M. Herlyn, and A. Raj, Nature **546**, 431 (2017).

[19] S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger, Nature **459**, 428 (2009).

[20] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).

[21] A. Raj and A. van Oudenaarden, Cell **135**, 216 (2008).

[22] D. Nicolas, B. Zoller, D. M. Suter, and F. Naef, Proc Natl Acad Sci U S A **115**, 7153 (2018).

[23] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, Science **332**, 472 (2011).

[24] M. R. Megaridis, Y. Y. Lu, E. N. Tevonian, K. M. Junger, J. M. Moy, K. Bohn-Wippert, and R. D. Dar, Apl Bioengineering **2** (2018), Artn 026106 10.1063/1.5021183.

[25] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger, Proc Natl Acad Sci U S A **109**, 17454 (2012).

[26] H. Einarsson, M. Salvatore, C. Vaagenso, N. Alcaraz, J. Lange, S. Rennie, and R. Andersson, eLife **11** (2022), 10.7554/eLife.80943.

[27] A. J. M. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, A. Segerstolpe, C. M. Rivera, B. Ren, and R. Sandberg, Nature **565**, 251 (2019).

[28] B. Zoller, D. Nicolas, N. Molina, and F. Naef, Molecular Systems Biology **11** (2015).

[29] V. L. Bass, V. C. Wong, M. E. Bullock, S. Gaudet, and K. Miller-Jensen, Mol Syst Biol **17**, e10127 (2021).

[30] V. C. Wong, V. L. Bass, M. E. Bullock, A. K. Chavali, R. E. C. Lee, W. Mothes, S. Gaudet, and K. Miller-Jensen, Cell Rep **22**, 585 (2018).

[31] N. Alachkar, D. Norton, Z. Wolkensdorfer, M. Muldoon, and P. Paszek, Frontiers in Molecular Biosciences **10** (2023), 10.3389/fmolb.2023.1176107.

[32] N. E. Phillips, A. Mandic, S. Omidi, F. Naef, and D. M. Suter, Nat Commun **10**, 1208 (2019).

[33] T. Ferraro, E. Esposito, L. Mancini, S. Ng, T. Lucas, M. Coppey, N. Dostatni, A. M. Walczak, M. Levine, and M. Lagha, Curr Biol **26**, 212 (2016).

[34] H. R. Clark, C. McKenney, N. M. Livingston, A. Gershman, S. Sajjan, I. S. Chan, A. J. Ewald, W. Timp, B. Wu, A. Singh, and S. Regot, Nature Communications **12**, 1836 (2021).

[35] J. J. Hughey, M. V. Gutschow, B. T. Bajar, and M. W. Covert, Mol Biol Cell **26**, 583 (2015).

[36] S. E. Luria and M. Delbruck, Genetics **28**, 491 (1943).

[37] C. R. Raetz, C. M. Reynolds, M. S. Trent, and R. E. Bishop, Annu Rev Biochem **76**, 295 (2007).

[38] K. V. Swanson, M. Deng, and J. P. Y. Ting, Nature Reviews Immunology **19**, 477 (2019).

[39] E. Ricciotti and G. A. FitzGerald, Arterioscler Thromb Vasc Biol **31**, 986 (2011).

[40] K. Oda and H. Kitano, Mol Syst Biol **2**, 2006 0015 (2006).

[41] K. Heyninck, D. De Valck, W. Vanden Berghe, W. Van Criekinge, R. Contreras, W. Fiers, G. Haegeman, and R. Beyaert, J Cell Biol **145**, 1471 (1999).

[42] E. G. Lee, D. L. Boone, S. Chai, S. L. Libby, M. Chien, J. P. Lodolce, and A. Ma, Science **289**, 2350 (2000).

[43] M. Son, A. G. Wang, H. L. Tu, M. O. Metzig, P. Patel, K. Husain, J. Lin, A. Murugan, A. Hoffmann, and S. Tay, Sci Signal **14** (2021), 10.1126/scisignal.aaz4382.

[44] A. V. Villarino, Y. Kanno, J. R. Ferdinand, and J. J. O'Shea, J Immunol **194**, 21 (2015).

[45] H. Feng, Y. B. Zhang, J. F. Gui, S. M. Lemon, and D. Yamane, PLoS Pathog **17**, e1009220 (2021).

[46] R. O. Hynes, Cell **110**, 673 (2002).

[47] J. M. Ankers, R. Awais, N. A. Jones, J. Boyd, S. Ryan, A. D. Adamson, C. V. Harper, L. Bridge, D. G. Spiller, D. A. Jackson, P. Paszek, V. See, and M. R. White, Elife **5** (2016), 10.7554/eLife.10473.

[48] I. N. Baranova, R. Kurlander, A. V. Bocharov, T. G. Vishnyakova, Z. Chen, A. T. Remaley, G. Csako, A. P. Patterson, and T. L. Eggerman, The Journal of Immunology **181**, 7147 (2008).

[49] D. Nicolas, N. E. Phillips, and F. Naef, Mol Biosyst **13**, 1280 (2017).

[50] V. Shahrezaei and P. S. Swain, Proc Natl Acad Sci U S A **105**, 17256 (2008).

[51] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, PLoS Biol **4**, e309 (2006).

[52] J. Moran, L. Feltham, J. Bagnall, M. C. Goldrick, E. A. Lord, C. Nettleton, D. G. Spiller, I. Roberts, and P. Paszek, bioRxiv , 2022.06.04.493993 (2022).

[53] D. A. C. Stapels, P. W. S. Hill, A. J. Westermann, R. A. Fisher, T. L. Thurston, A. E. Saliba, I. Blommestein, J. Vogel, and S. Helaine, Science **362**, 1156 (2018).

[54] L. Ashall, C. A. Horton, D. E. Nelson, P. Paszek, C. V. Harper, K. Sillitoe, S. Ryan, D. G. Spiller, J. F. Unitt, D. S. Broomhead, D. B. Kell, D. A. Rand, V. See, and M. R. White, Science **324**, 242 (2009).

[55] T. Lipniacki, P. Paszek, A. R. Brasier, B. A. Luxon, and M. Kimmel, Biophys J **90**, 725 (2006).

[56] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, Molecular Cell **58**, 339 (2015).

[57] N. Battich, T. Stoeger, and L. Pelkmans, Cell **163**, 1596 (2015).

[58] L. C. Davies, M. Rosas, S. J. Jenkins, C. T. Liao, M. J. Scurr, F. Brombacher, D. J. Fraser, J. E. Allen, S. A. Jones, and P. R. Taylor, Nat Commun **4**, 1886 (2013).

[59] E. R. Sherwood, K. R. Burelbach, M. A. McBride, C. L. Stothers, A. M. Owen, A. Hernandez, N. K. Patil, D. L. Williams, and J. K. Bohannon, J Immunol **208**, 785 (2022).

[60] E. M. Jeff, H. W. Martin, R. Michael, H.-J. Michael, H. Joanna, E. Carl-Johan, T. Hosein, B. Joseph, B. Leonie von, M. Marcel, B. Kim, L. Jens, L. Joakim, S. Rickard, M. Jakob, and F. Jonas, bioRxiv , 2022.02.14.480352 (2022).

[61] Z. Meir, Z. Mukamel, E. Chomsky, A. Lifshitz, and A. Tanay, Nature Genetics **52**, 709 (2020).

[62] T. Muramoto, I. Müller, G. Thomas, A. Melvin, and J. R. Chubb, Curr Biol **20**, 397 (2010).

[63] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, Cell Rep **15**, 2038 (2016).

[64] V. Hornung, F. Bauernfeind, A. Halle, E. O. Samstad, H. Kono, K. L. Rock, K. A. Fitzgerald, and E. Latz, Nature Immunology **9**, 847 (2008).

[65] A. M. Bolger, M. Lohse, and B. Usadel, Bioinformatics **30**, 2114 (2014).

[66] W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler, Genome Res **12**, 996 (2002).

[67] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras, Bioinformatics **29**, 15 (2013).

[68] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, and T. J. Hubbard, Genome Res **22**, 1760 (2012).

[69] M. I. Love, W. Huber, and S. Anders, Genome Biology **15**, 550 (2014).
[70] A. Zhu, J. G. Ibrahim, and M. I. Love, Bioinformatics **35**, 2084 (2018).
[71] R. A. Amezquita, A. T. L. Lun, E. Becht, V. J. Carey, L. N. Carpp, L. Geistlinger, F. Marini, K. Rue-Albrecht, D. Risso, C. Soneson, L. Waldron, H. Pages, M. L. Smith, W. Huber, M. Morgan, R. Gottardo, and S. C. Hicks, Nat Methods **17**, 137 (2020).
[72] Z. Xie, A. Bailey, M. V. Kuleshov, D. J. B. Clarke, J. E. Evangelista, S. L. Jenkins, A. Lachmann, M. L. Wojciechowicz, E. Kropiwnicki, K. M. Jagodnik, M. Jeon, and A. Ma'ayan, Current Protocols **1**, e90 (2021).
[73] M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, and A. Ma'ayan, Nucleic Acids Res **44**, W90 (2016).

Supplementary materials

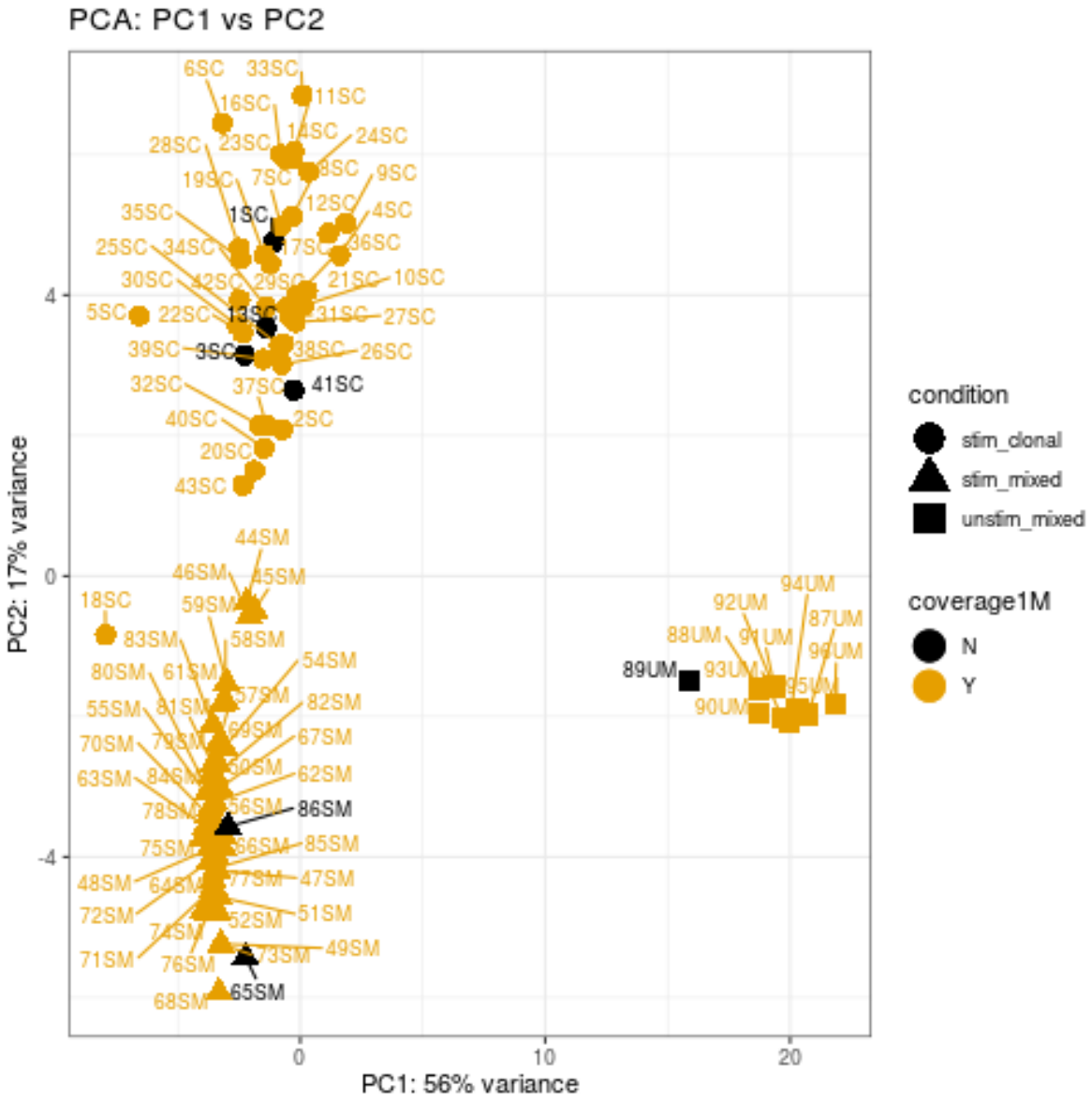


Figure supplement 1: Analysis of MemorySeq cell populations. Shown is a PCA analyses of stimulated 42 clonal (SC, stim_clonal), 43 mixed (SM, stim_mixed) and unstimulated mixed population (UM, unstim_mixed). Variance explained by each principal component highlighter in the axis label. In black samples with a coverage between 0.5 and 1 million reads. Sample 18SC was removed from subsequent analyses.

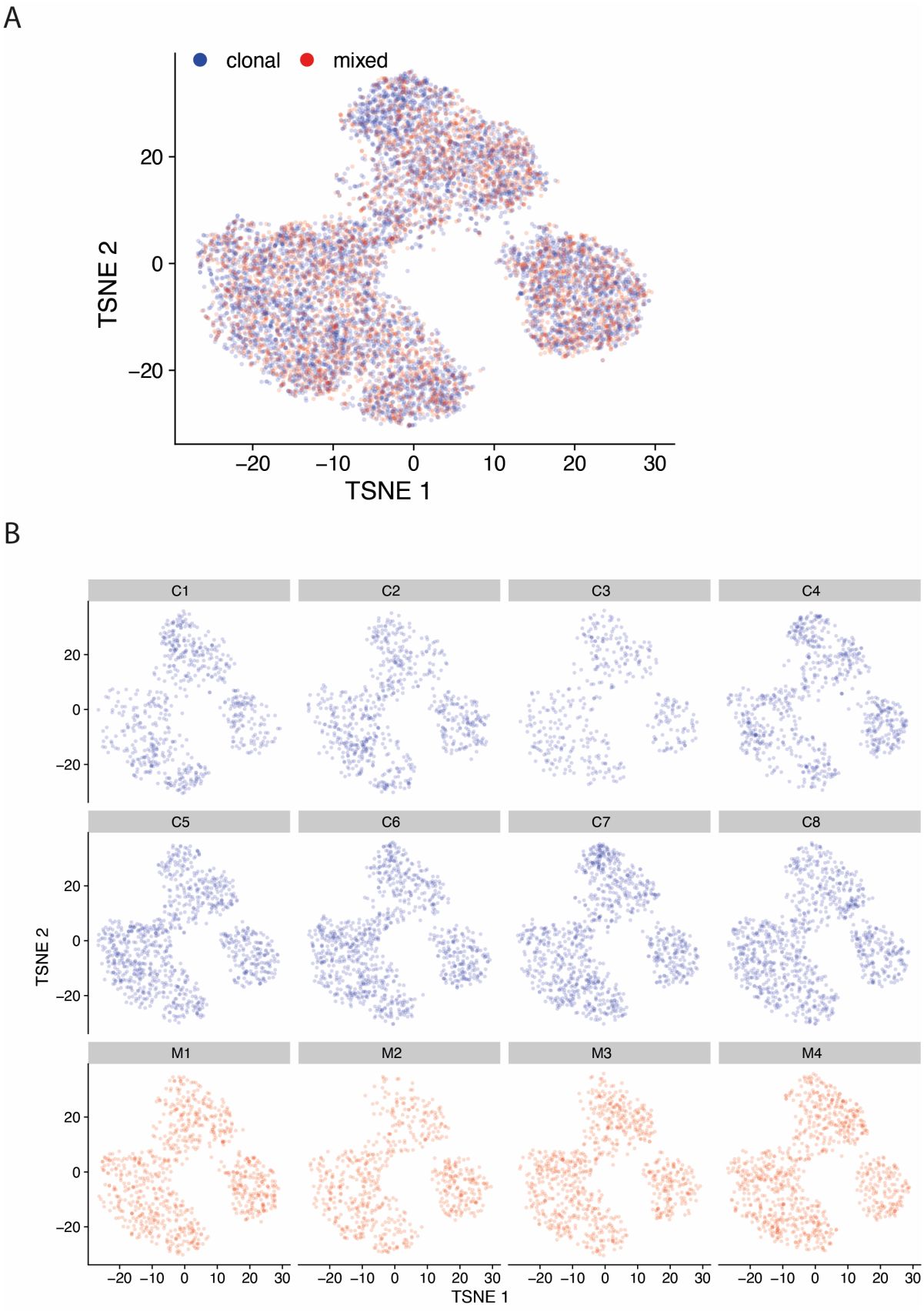


Figure supplement 2: Analysis of the scRNA-seq populations. A. t-SNE plot for clonal and mixed cells grouped together. **B.** t-SNE plots of individual clonal (C1 to C8) and mixed (M1 to M4) cells.

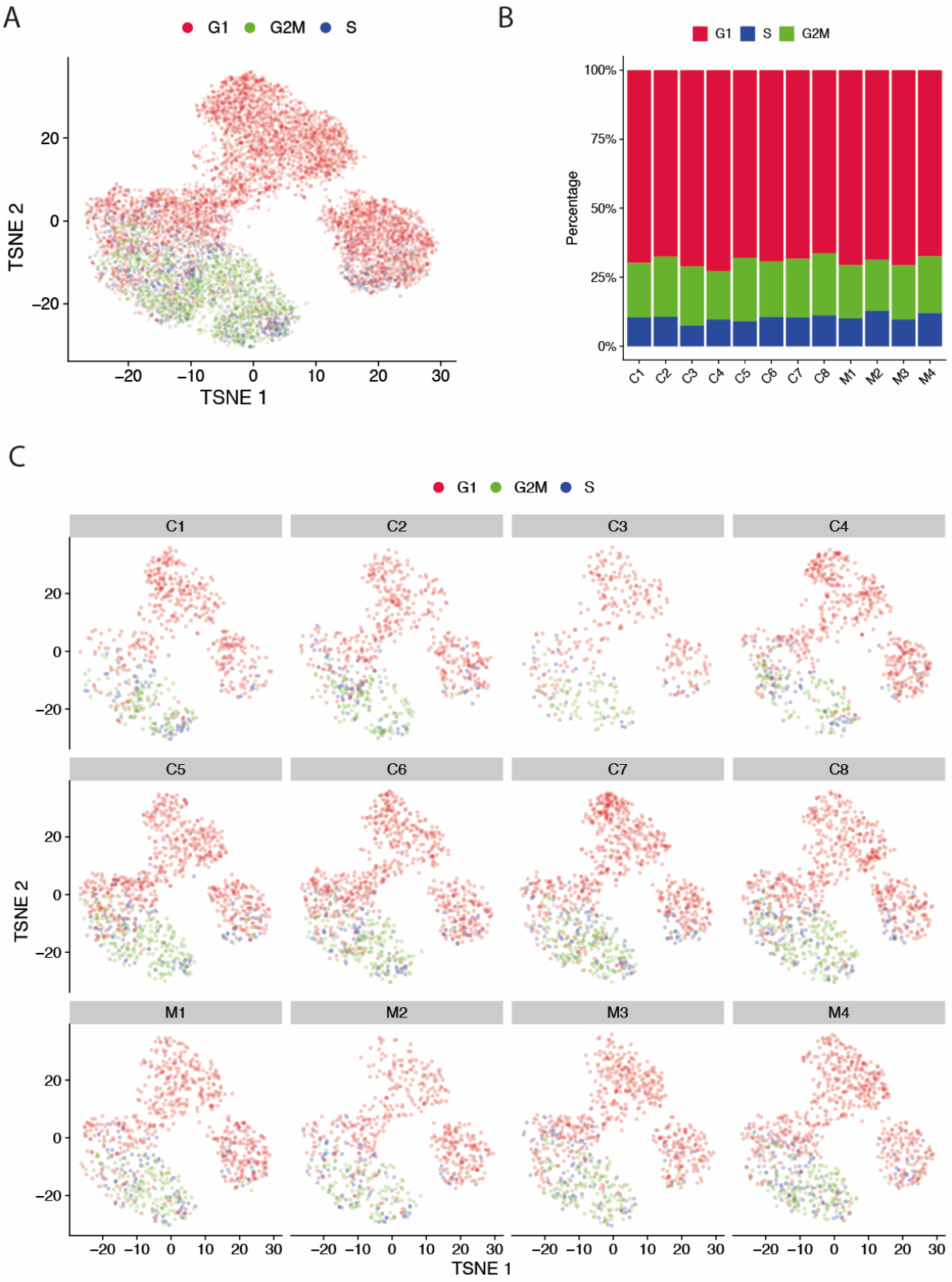


Figure supplement 3: Cell cycle analysis of the scRNA-seq populations. A. t-SNE plot for clonal and mixed cells grouped together, with the inferred cell cycle stage superimposed on to the graph (in different colours). **B.** Distribution of cell cycle stages across individual clonal (C1 to C8) and mixed (M1 to M4) cells. **C.** t-SNE plots of individual clonal (C1 to C8) and mixed (M1 to M4) cells, with the corresponding cell cycle stage.

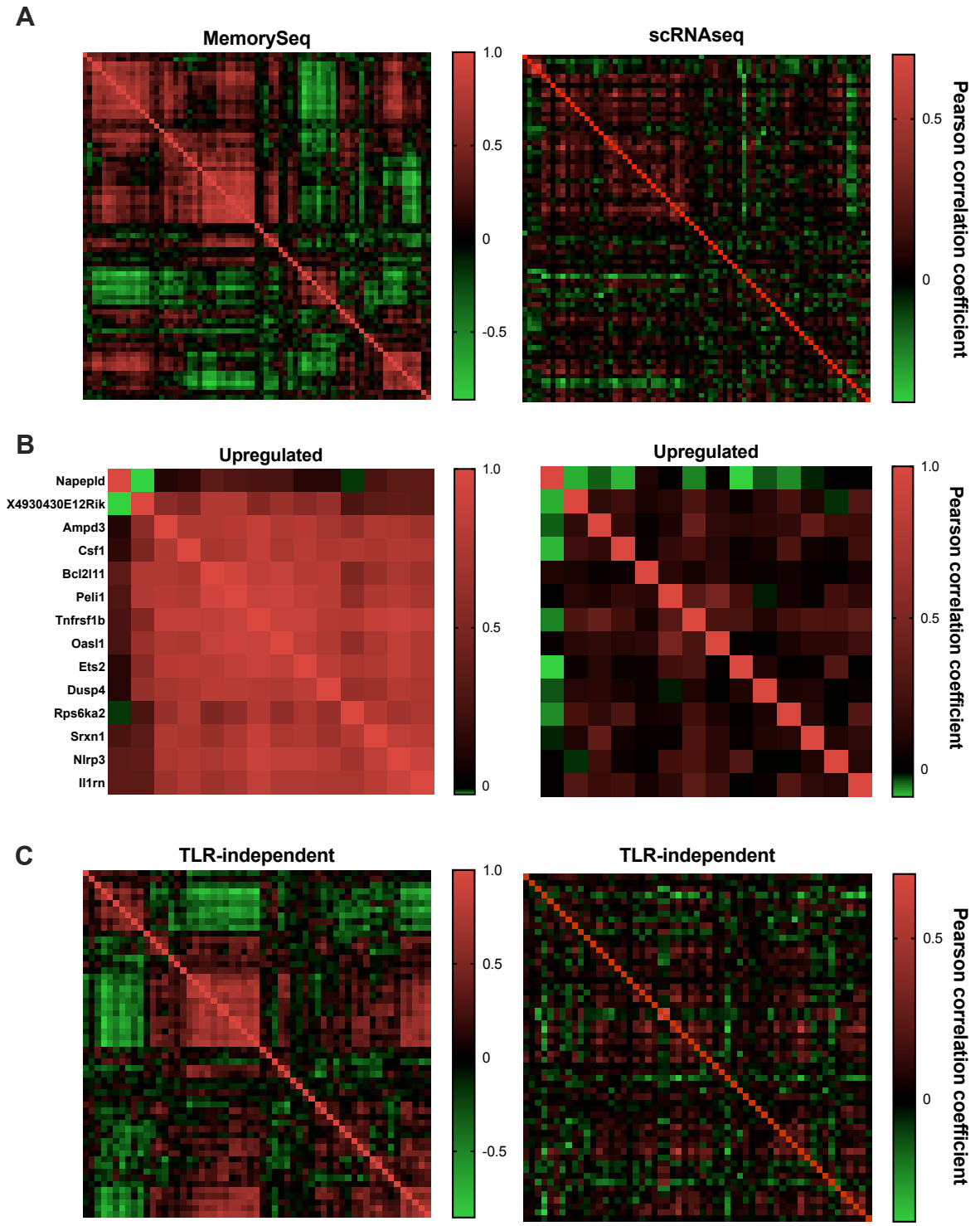


Figure supplement 4: Heritable genes exhibit co-fluctuations. Shown are Pearsons's correlation heatmaps for all pairs of 73 heritable genes in both MemorySeq (left) and scRNA-seq datasets (right). **A.** Correlation heatmap of all heritable genes. **B.** Correlation heatmaps of up-regulated heritable genes. **C.** Correlation heatmaps of TLR-independent heritable genes. Heatmaps for scRNA-seq data displayed with ordering of the MemorySeq correlation matrix.

Chapter 5 Discussion

In this thesis, I have developed novel mathematical methodologies, led by experimental data, to quantitatively characterise transcriptional heterogeneity and the behaviour of its underlying bursting characteristics in the TLR system. Through this work, various stochastic models of gene expression were formulated, different experimental techniques and statistical analysis tools were used, and new concepts regarding heritability of variability were introduced. All of which contributed to a deeper understanding of the complex nature of gene expression and its variability and provided new perspectives for its quantitative description.
In this chapter I outline how the aims set forth in the introduction have been successfully accomplished, and discuss limitations of the work, in addition to potential avenues for future research in the field.

The aims were achieved as follows:

Characterisation of transcriptional heterogeneity in the TLR system

Transcriptional responses of two important immune genes, *TNF*α and *IL*1β, were examined under immune-relevant stimulus using smFISH data. The analysis revealed different levels of cell-to-cell variability for these two genes, with *IL*1β exhibiting higher variability than *TNF*α across conditions. To further explore their regulatory differences, I developed two stochastic models with varying levels of complexity and showed that mRNA counts of *TNF*α conform to a standard stochastic switch model, while transcription of *IL*1β requires an additional regulatory step reflecting increased heterogeneity. This suggested that heterogeneity can be gene-specific and controlled by certain transcriptional mechanisms. To test this hypothesis, I checked for and established gene-specific linear relationship between the mean and variance of *TNF*α and *IL1*β mRNA responses to the different conditions, reflecting a fundamental characteristic of the gene regulatory system (Dar et al., 2016). This was confirmed further for more genes using scRNA-seq data. I also showed that this constrained variability arises from constrained modulation of transcriptional bursting. Under the assumption of a two-state model, I derived an analytical function that describes a relationship between model parameters that follows from the mean-variance linear function, defining a three-dimensional (k_{off}, k_{on}, k_t) parameter space (assuming k_d is known) on which the system can maintain the mean-variance relationship by constrained modulation in both burst size and

frequency kinetics. I verified my results by examining *TNF*α and *IL*1β bursting characteristics and their kinetics using smFISH data.

More questions for consideration emerged from this study; are these findings global if more immune genes are considered? Are the linear constraints limited to pathogen-like stimulus, and specifically the set of conditions used in our study (Oda & Kitano, 2006b), or do they hold under therapeutic compounds conditions too? The failure of the two-state model to adequately reflect the high variability of *IL*1β expression suggested that the level of transcriptional variability is related to the complexity of transcriptional regulation mechanisms. What are these complex molecular mechanisms resulting in high levels of variability? And do genes that exhibit similar variability levels share the same complex regulatory modes? Chapter 3 addresses some of these questions while further studies are required to link some already existing knowledge about regulatory mechanisms (Junkin et al., 2016; Shalek et al., 2013a; Singer et al., 2014; Xue et al., 2015).

Nevertheless, this chapter provided evidence that seemingly heterogeneous gene expression is controlled and provided a mathematical framework to accurately characterise the underlying mechanisms driving this heterogeneity. In addition, due to the importance of *TN*Fα and *IL*1β in regulating the innate immune system, the dynamical models established for them in this thesis can be utilized in other immune research studies.

Characterisation of transcriptional bursting modulation in the TLR system

Utilising existing scRNA-seq data, I investigated whether the control of transcriptional bursting in the TLR system is a global property. Analysis of > 2000 TLR-response genes across different experimental conditions revealed that 66% of the genes show gene-specific invariant cellular variability levels across the different conditions, i.e., the changes in sample mean of these genes due to change in stimulus yielded to equal proportional changes in sample variance. This result confirmed that characterising variability in the TLR system by linear mean-variance trends using linear regression models is an effective approach to understand constraints of cellular variability. The analysis further confirmed that gene-specific response variability is a global trait, identified by the slope of the mean-variance relationship. Next, I explored the properties of transcriptional bursting underlying the global

gene-specific mean-variance trends. I derived from the linear relationship analytical functions describing burst size and frequency changes with changes in mean mRNA response based on their definitions using moment estimators (mean and variance). The derived functions tell us that burst size is a multiplicative inverse of the mean mRNA, converging to the slope value as mean increases, which is why in some cases burst size can be used as a measure of variability level, and that burst frequency converges quickly to the mean scaled by the inverse of the slope. Experimental data showed strong agreement with the theory, validating the theoretical results further.

Although I present burst size and frequency as functions of the mean expression, biologically the mean is the result of both the burst size and the frequency dynamics. Therefore, I sought to study the behaviour of burst size and frequency independently of the mean. Deriving frequency behaviour as a function of burst size uncovered three different regulatory modes, with overall modulation in the system being dominated by frequency changes. Analysis of the experimental data confirmed the defined regulatory modes. This suggests that the control of variability is a result of an interplay between different molecular regulatory factors including core promoters (which have been established to control burst size modulation), and enhancer elements (which are associated with the control of frequency modulation) (Larsson et al., 2019). It would be intriguing to verify this on the genes considered in this study and establish regulatory mechanisms and promoter properties that shape the gene-specific slope and hence gene expression patterns. Once this is done, the results can be incorporated into the mathematical model to deliver a more complete picture of the process.

In addition, the analysed genes exhibited a broad spectrum of variability levels, as evidenced by the diverse range of slope values observed. This, combined with the findings from chapter two that employed different stochastic models to describe TNF-α and IL-1β, suggested that the mechanistic control of variability varies depending on its magnitude. I hypothesised that a high level of variability is associated with a heightened degree of transcriptional regulation complexity.

To test this hypothesis, I utilised a 3-state stochastic model that incorporates sequential promoter activation, akin to promoter cycling (Zoller et al., 2015), in addition to the simple 2-state model. I fitted a set of 96 genes and determined which model fits best for each condition of the genes using AIC. I found that not only were the bursting characteristics of conditions

better fitted with the 3-state model significantly different from those better fitted with the 2-state model, but there was also a positive correlation between the probability of a gene conforming to a 3-state model and an increase in slope, confirming my hypothesis.
I also investigated the control of variability through evolution and provided evidence that mean-variance linear trends serve as a valuable tool to capture and understand the evolutionarily changes in response variability.

One challenge I encountered during the modelling phase of the research was selecting the optimal method for parameters estimation. The genetic algorithm employed in this study has the potential drawback of premature convergence, where the algorithm converges too quickly without fully exploring the entire parameter space, especially in the case of a more complex model, leading to a poor fit. A Bayesian inference approach remains the most robust and reliable method to infer model parameter. While tools for inferring parameters from scRNA-seq data using a Bayesian inference framework are currently limited, there is promising ongoing work in this area (Breda et al., 2021).

One potential avenue for improving this research is to explore models that incorporate time-dependent parameters. By allowing parameters to vary over time (Dattani & Barahona, 2017), we can gain a more comprehensive understanding of the regulation of cellular variability upon perturbations, providing a more accurate representation of the underlying biological processes. It is reasonable to say, however, that with this approach, computational challenges would arise, and much more accurate and reliable dynamic data would be required.

Heterogeneity observed in the innate immune system is a heritable trait – challenging the dogma of stochastically regulated transcription

Using bulk and single-cell RNA-seq datasets, I uncovered heritability of single-cell expression patterns, particularly rare expression events, in the TLR system, presenting a novel perspective on the stochastic nature of gene expression. This finding challenges the commonly used stochastic model of gene expression, the telegraph model, and opens new avenues for understanding the intricacies of gene regulation and thereby heterogeneity control mechanisms.

Analysis of MemorySeq data unveiled the intriguing phenomenon that certain TLR-dependent genes possess the capability to pass on their non-genetic expression fluctuations across multiple generations through cell division. This was further supported by additional analyses using complementary scRNA-seq and immunostaining approaches. This work aligns with the few recent studies that provided evidence of transcriptional memory for responses of rare cells in different cell lines to the ones used here (Phillips et al., 2019; Shaffer et al., 2020).
Subsequently, I proposed the hypothesis that the differences observed between clonal populations of heritable genes are indicative of variations in clone-specific transcriptional control. To test this hypothesis, I investigated whether there are similarities in the transcriptional characteristics among clonal populations of heritable genes. Based on the framework of the mean-variance relationship, the noise characteristics observed in clonal populations were found to deviate from each other and those of the parental cells. The parental cells were represented as a single point on the gene's mean-variance line, subject to technical noise. However, the clonal populations displayed a divergence from this point and a dispersion within the clones, while still conforming to a linear model. It follows from my previous work in chapters two and three (Alachkar et al., 2023; Bagnall et al., 2020) that they have clone-specific predictive bursting characteristics, indicating variations in the mechanisms controlling heterogeneity among different clones. The defined measure of heritability in the study indicates that this trait is not simply binary but instead transiently dynamic, as reflected in genes exhibiting a spectrum of heritability levels, which are most likely contingent upon the timescale of cell division. Therefore, I predict that once genes pass the timescale of transcriptional memory, they converge back towards the parental populations on the mean-variance linear relationship while showing similar bursting characteristics. Supporting evidence of this came from the analysis of non-heritable genes, where clonal populations displayed comparable noise characteristics to the parental populations.

Analysis of transcriptional bursting further revealed that the primary determinant of the differences observed among individual clonal populations is frequency modulation rather than burst size modulation. Different controlling mechanisms of frequency modulation have been established, including epigenetic factors (Larsson et al., 2019; Nicolas et al., 2018). In particular, frequency modulation associated with single-cell TLR-dependent expression has

been shown to be controlled through cohesin-mediated promoter-enhancer coupling (Robles-Rebollo et al., 2022). On the other hand, transient heritability has been demonstrated to be associated with long-term epigenetic control (Van Eyndhoven et al., 2023). Investigating the findings of these studies experimentally on the heritable genes identified here would contribute to our understanding of the underlying mechanisms behind the regulation of heterogeneity and its heritability trait.

The dynamic aspect of the heritability trait remains unexplored here, and an important future addition to this work would involve the use of real-time RNA imaging techniques, such as CRISPR-Cas13 systems (Yang et al., 2019), in live cells. By investigating the dynamical behaviour of heritability across several generations of dividing cells, not only would we gain valuable insights into the regulatory mechanisms underlying cellular variability at a temporal level, but also refine our understanding of the control of immune responses in the TLR system. Furthermore, the findings of this study highlight the necessity for a different mathematical modelling approach that goes beyond the currently accepted stochastic models in the field. It is evident that there are deterministic factors involved in the regulation of heterogeneity, and these factors should be incorporated into the dynamical models of gene expression. Recent studies presented some tools regarding this task (Saint-Antoine et al., 2022; Van Eyndhoven et al., 2023) by providing a model that accounts for reversable switching between cell-states (defined by responsive or not). Improving these tools to account for different cell states (representing different heritability levels) and integrating them with data from dynamic imaging to infer the switching rates would ultimately present a more comprehensive model that captures a closer approximation of the biological phenomenon.
Overall, this study has provided a novel perspective on transcriptional heterogeneity, and paved the way for future research, both in the biological and mathematical domains.

In conclusion, the findings presented in this thesis strongly support the speculated concept that cellular variability plays a crucial role in immune responses, particularly during inflammation. The newfound understanding of the underlying mechanisms of this variability presented here does not only advance our knowledge of gene expression in general but also provides deeper insights into the functioning features of the immune system. These results have the potential to be integrated into existing models of the TLR system and, more broadly,

the innate immune system. By incorporating this knowledge, we can enhance our understanding of immune responses and potentially develop more effective strategies for therapeutic interventions in various immune-related disorders.

Overall, while this work has made significant contributions to our understanding of transcriptional heterogeneity and its regulation, there is still much to be explored in the field.

References

Adamik, J., Wang, K. Z. Q., Unlu, S., Su, A. J. A., Tannahill, G. M., Galson, D. L., O'Neill, L. A., & Auron, P. E. (2013). *Distinct Mechanisms for Induction and Tolerance Regulate the Immediate Early Genes Encoding Interleukin 1 beta and Tumor Necrosis Factor alpha*.

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen Recognition and Innate Immunity. *Cell*, *124*(4), 783–801. https://doi.org/10.1016/j.cell.2006.02.015

Alachkar, N., Norton, D., Wolkensdorfer, Z., Muldoon, M., & Paszek, P. (2023). Variability of the innate immune response is globally constrained by transcriptional bursting. *Frontiers in Molecular Biosciences*, *10*. https://doi.org/10.3389/fmolb.2023.1176107

Al-Mohy, A. H., & Higham, N. J. (2011). Computing the Action of the Matrix Exponential, with an Application to Exponential Integrators. *SIAM Journal on Scientific Computing*, *33*(2), 488–511. https://doi.org/10.1137/100788860

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Nature Precedings*. https://doi.org/10.1038/npre.2010.4282.2

Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., François, J., & Yvert, G. (2008). Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet*, *4*, 1000049.

Antczak, M., Cañete, P. F., Chen, Z., Belle, C., & Yu, D. (2022). Evolution of γ chain cytokines: Mechanisms, methods and applications. *Computational and Structural Biotechnology Journal*, *20*, 4746–4755. https://doi.org/10.1016/j.csbj.2022.08.050

Arango Duque, G., & Descoteaux, A. (2014). Macrophage Cytokines: Involvement in Immunity and Infectious Diseases. *Frontiers in Immunology*, *5*, 491. https://doi.org/10.3389/fimmu.2014.00491

Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A., & Blom, J. G. (2009). Systems biology: Parameter estimation for biochemical models. *The FEBS Journal*, *276*(4), 886–902. https://doi.org/10.1111/j.1742-4658.2008.06844.x

Ashyraliyev, M., Jaeger, J., & Blom, J. G. (2008). Parameter estimation and determinability analysis applied to Drosophila gap gene circuits. *BMC Systems Biology*, *2*, 83.

Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H. B., Trombetta, J. J., Satija, R., Shalek, A. K., Xavier, R. J., Regev, A., & Hung, D. T. (2015). Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell*, *162*(6), 1309–1321. https://doi.org/10.1016/j.cell.2015.08.027

Bagnall, J., Boddington, C., England, H., Brignall, R., Downton, P., Alsoufi, Z., Boyd, J., Rowe, W., Bennett, A., & Walker, C. (2018). *Quantitative analysis of competitive cytokine signaling predicts tissue thresholds for the propagation of macrophage activation*.

Bagnall, J., Rowe, W., Alachkar, N., Roberts, J., England, H., Clark, C., Platt, M., Jackson, D. A., Muldoon, M., & Paszek, P. (2020). Gene-Specific Linear Trends Constrain Transcriptional Variability of the Toll-like Receptor Signaling. *Cell Systems*, *11*(3), 300-314.e8. https://doi.org/10.1016/j.cels.2020.08.007

Bahar Halpern, K., & Itzkovitz, S. (2016). Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues. *Methods (San Diego, Calif*, *98*, 134–142.

Bahar Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A., & Itzkovitz, S. (2015). Bursty gene expression in the intact mammalian liver. *Molecular Cell*, *58*, 147–156.

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., & Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science*, *305*, 1622–1625.

Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A., & Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular Cell*, *62*, 237–247.

Batenchuk, C., St-Pierre, S., Tepliakova, L., Adiga, S., Szuto, A., Kabbani, N., Bell, J. C., Baetz, K., & Kærn, M. (2011). Chromosomal position effects are linked to sir2-mediated variation in transcriptional burst size. *Biophysical Journal*, *100*, 56–58.

Beckman, W. F., Jiménez, M. Á. L., & Verschure, P. J. (2021). Transcription bursting and epigenetic plasticity: An updated view. *Epigenetics Communications*, *1*(1), 6. https://doi.org/10.1186/s43682-021-00007-1

Becskei, A., Kaufmann, B. B., & van Oudenaarden, A. (2005). Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genetics*, *37*(9), Article 9. https://doi.org/10.1038/ng1616

Bengtsson, M., Ståhlberg, A., Rorsman, P., & Kubista, M. (2005). Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research*, *15*(10), 1388–1392. https://doi.org/10.1101/gr.3820805

Bix, M., & Locksley, R. M. (1998). Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. *Science (New York, N.Y.)*, *281*(5381), 1352–1354. https://doi.org/10.1126/science.281.5381.1352

Blake, W. J., Balazsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., Cantor, C. R., Walt, D. R., & Collins, J. J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular Cell*, *24*, 853–865.

Blake, W. J., KÆrn, M., Cantor, C. R., & Collins, J. J. (2003a). Noise in eukaryotic gene expression. *Nature*, *422*, 633–637.

Blake, W. J., KÆrn, M., Cantor, C. R., & Collins, J. J. (2003b). Noise in eukaryotic gene expression. *Nature*, *422*(6932), Article 6932. https://doi.org/10.1038/nature01546

Breda, J., Zavolan, M., & van Nimwegen, E. (2021). Bayesian inference of gene expression states from single-cell RNA-seq data. *Nature Biotechnology*, *39*(8), Article 8. https://doi.org/10.1038/s41587-021-00875-x

Brown, C. R., & Boeger, H. (2014). Nucleosomal promoter variation generates gene expression noise. *Proceedings of the National Academy of Sciences*, *111*(50), 17893–17898. https://doi.org/10.1073/pnas.1417527111

Bryant, C. E., Symmons, M., & Gay, N. J. (2015). Toll-like receptor signalling through macromolecular protein complexes. *Mol Immunol*, *63*, 162–165.

Cao, H., Wang, Y., Zhang, N., Xia, S., Tian, P., Lu, L., Du, J., & Du, Y. (2022). Progress of CRISPR-Cas13 Mediated Live-Cell RNA Imaging and Detection of RNA-Protein Interactions. *Frontiers in Cell and Developmental Biology*, *10*, 866820. https://doi.org/10.3389/fcell.2022.866820

Cao, Y., Gillespie, D. T., & Petzold, L. R. (2006). Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics*, *124*, 044109.

Cao, Y., Li, H., & Petzold, L. (2004). Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J Chem Phys*, *121*, 4059–4067.

Cao, Z., & Grima, R. (2020). Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proceedings of the National Academy of Sciences*, *117*(9), 4682–4692. https://doi.org/10.1073/pnas.1910888117

Capp, J.-P. (2021). Interplay between genetic, epigenetic, and gene expression variability: Considering complexity in evolvability. *Evolutionary Applications*, *14*(4), 893–901. https://doi.org/10.1111/eva.13204

Cesbron, F., Oehler, M., Ha, N., Sancar, G., & Brunner, M. (2015). Transcriptional refractoriness is dependent on core promoter architecture. *Nature Communications*, *6*(1), Article 1. https://doi.org/10.1038/ncomms7753

Chaplin, D. D. (2010). Overview of the Immune Response. *The Journal of Allergy and Clinical Immunology*, *125*(2 Suppl 2), S3-23. https://doi.org/10.1016/j.jaci.2009.12.980

Chen, C.-Y. A., Ezzeddine, N., & Shyu, A.-B. (2008). Messenger RNA Half-Life Measurements in Mammalian Cells. *Methods in Enzymology*, *448*, 335–357. https://doi.org/10.1016/S0076-6879(08)02617-7

Chen, J.-Y., Lin, J.-R., Cimprich, K. A., & Meyer, T. (2012). A two-dimensional ERK-AKT signaling code for an NGF-triggered cell-fate decision. *Molecular Cell*, *45*(2), 196–209. https://doi.org/10.1016/j.molcel.2011.11.023

CHEN, Y., LIN, J., ZHAO, Y., MA, X., & YI, H. (2021). Toll-like receptor 3 (TLR3) regulation mechanisms and roles in antiviral innate immune responses. *Journal of Zhejiang University. Science. B*, *22*(8), 609–632. https://doi.org/10.1631/jzus.B2000808

Chubb, J. R., Trcek, T., Shenoy, S. M., & Singer, R. H. (2006). Transcriptional pulsing of a developmental gene. *Current Biology: CB*, *16*(10), 1018–1025. https://doi.org/10.1016/j.cub.2006.03.092

Cianfrocco, M. A., Kassavetis, G. A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J. T., & Nogales, E. (2013). Human TFIID Binds to Core Promoter DNA in a Reorganized Structural State. *Cell*, *152*(1), 120–131. https://doi.org/10.1016/j.cell.2012.12.005

Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V., McCollum, J. M., Cox, C. D., Simpson, M. L., & Weinberger, L. S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences*, *109*, 17454–17459.

Dar, R. D., Shaffer, S. M., Singh, A., Razooky, B. S., Simpson, M. L., Raj, A., & Weinberger, L. S. (2016). Transcriptional Bursting Explains the Noise–Versus–Mean Relationship in mRNA and Protein Levels. *PLOS ONE*, *11*, 0158298.

Dattani, J., & Barahona, M. (2017). Stochastic models of gene transcription with upstream drives: Exact solution and sample path characterization. *Journal of The Royal Society Interface*, *14*(126), 20160833. https://doi.org/10.1098/rsif.2016.0833

de Jonge, W. J., O'Duibhir, E., Lijnzaad, P., van Leenen, D., Groot Koerkamp, M. J., Kemmeren, P., & Holstege, F. C. (2017). Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *The EMBO Journal*, *36*(3), 274–290. https://doi.org/10.15252/embj.201695621

de Torrenté, L., Zimmerman, S., Suzuki, M., Christopeit, M., Greally, J. M., & Mar, J. C. (2020). The shape of gene expression distributions matter: How incorporating distribution shape improves the interpretation of cancer transcriptomic data. *BMC Bioinformatics*, *21*(21), 562. https://doi.org/10.1186/s12859-020-03892-w

Dey, S. S., Foley, J. E., Limsirichai, P., Schaffer, D. V., & Arkin, A. P. (2015). Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Molecular Systems Biology*, *11*, 806.

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, *403*, 335–338.

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic Gene Expression in a Single Cell. *Science*, *297*, 1183–1186.

Escoubet-Lozach, L., Benner, C., Kaikkonen, M. U., Lozach, J., Heinz, S., Spann, N. J., Crotti, A., Stender, J., Ghisletti, S., & Reichart, D. (2011). Mechanisms establishing TLR4-

responsive activation states of inflammatory response genes. *PLoS Genet*, *7*, 1002401.

Ezer, D., Moignard, V., Göttgens, B., & Adryan, B. (2016). Determining Physical Mechanisms of Gene Expression Regulation from Single Cell Gene Expression Data. *PLOS Computational Biology*, *12*, 1005072.

Femino, A. M., Fay, F. S., Fogarty, K., & Singer, R. H. (1998). Visualization of single RNA transcripts in situ. *Science*, *280*, 585–590.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I. K., Sharon, E., Lubling, Y., Widom, J., & Segal, E. (2008). Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals. *PLOS Computational Biology*, *4*, 1000216.

Foreman, R., & Wollman, R. (2020). Mammalian gene expression variability is explained by underlying cell state. *Molecular Systems Biology*, *16*(2), e9146. https://doi.org/10.15252/msb.20199146

Fraser, D., & Kaern, M. (2009). A chance at survival: Gene expression noise and phenotypic diversification strategies. *Molecular Microbiology*, *71*, 1333–1340.

Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, *166*, 358–368.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. In *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6* (pp. 721–741).

Genetic Algorithms (pp. 91–100). (2021). Academic Press. https://doi.org/10.1016/B978-0-12-821986-7.00013-5

Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, *105*(1), Article 1. https://doi.org/10.1038/hdy.2010.54

Gibson, M. A., & Bruck, J. (2000). Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A*, *104*, 1876–1889.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, *81*, 2340–2361.

Gillespie, D. T. (1992). A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and Its Applications*, *188*, 404–425.

Gillespie, D. T. (2007). Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry*, *58*, 35–55.

Girolami, M. (n.d.). *A first course in machine learning*.

Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, *123*, 1025–1036.

Gómez-Schiavon, M., Chen, L.-F., West, A. E., & Buchler, N. E. (2017). BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. *Genome Biology*, *18*, 164.

Greenberg, J. R. (1972). High Stability of Messenger RNA in Growing Cultured Cells. *Nature*, *240*(5376), Article 5376. https://doi.org/10.1038/240102a0

Grote, J., Krysciak, D., & Streit, W. R. (2015). Phenotypic Heterogeneity, a Phenomenon That May Explain Why Quorum Sensing Does Not Always Result in Truly Homogenous Cell Behavior. *Applied and Environmental Microbiology*, *81*(16), 5280–5289. https://doi.org/10.1128/AEM.00900-15

Guo, G., Pinello, L., Han, X., Lai, S., Shen, L., Lin, T.-W., Zou, K., Yuan, G.-C., & Orkin, S. H. (2016). Serum-Based Culture Conditions Provoke Gene Expression Variability in

Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis. *Cell Reports*, *14*(4), 956–965. https://doi.org/10.1016/j.celrep.2015.12.089

Guo, L., Hu-Li, J., & Paul, W. E. (2005). Probabilistic regulation of IL-4 production. *Journal of Clinical Immunology*, *25*(6), 573–581. https://doi.org/10.1007/s10875-005-8218-5

Hagai, T., Chen, X., Miragaia, R. J., Rostom, R., Gomes, T., Kunowska, N., Henriksson, J., Park, J.-E., Proserpio, V., Donati, G., Bossini-Castillo, L., Vieira Braga, F. A., Naamati, G., Fletcher, J., Stephenson, E., Vegh, P., Trynka, G., Kondova, I., Dennis, M., … Teichmann, S. A. (2018). Gene expression variability across cells and species shapes innate immunity. *Nature*, *563*(7730), 197–202. https://doi.org/10.1038/s41586-018-0657-2

Handy, D. E., Castro, R., & Loscalzo, J. (2011). Epigenetic Modifications: Basic Mechanisms and Role in Cardiovascular Disease. *Circulation*, *123*(19), 2145–2156. https://doi.org/10.1161/CIRCULATIONAHA.110.956839

Hao, S., & Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nature Immunology*, *10*(3), Article 3. https://doi.org/10.1038/ni.1699

Hargrove, J. L., Hulsey, M. G., & Beale, E. G. (1991). The kinetics of mammalian gene expression. *BioEssays*, *13*(12), 667–674. https://doi.org/10.1002/bies.950131209

Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D. G., Mullins, J. J., Rand, D. A., & Davis, J. R. E. (2011). Dynamic Analysis of Stochastic Transcription Cycles. *PLOS Biology*, *9*, 1000607.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, *57*, 97–109.

Hayden, M. S., West, A. P., & Ghosh, S. (2006). NF-κB and the immune response. *Oncogene*, *25*(51), Article 51. https://doi.org/10.1038/sj.onc.1209943

Haynes, W. (2013). Maximum Likelihood Estimation. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 1190–1191). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_1235

Hooshangi, S., Thiberge, S., & Weiss, R. (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(10), 3581–3586. https://doi.org/10.1073/pnas.0408507102

Horsthemke, W., Doering, C. R., Ray, T. S., & Burschka, M. A. (1992). Fluctuations and correlations in a diffusion-reaction system: Unified description of internal fluctuations and external noise. *Physical Review A*, *45*, 5492–5503.

Hu, J., Nudelman, G., Shimoni, Y., Kumar, M., Ding, Y., López, C., Hayot, F., Wetmur, J. G., & Sealfon, S. C. (2011). Role of Cell-to-Cell Variability in Activating a Positive Feedback Antiviral Response in Human Dendritic Cells. *PLOS ONE*, *6*, 16614.

Huang, S. (2009). Non-genetic heterogeneity of cells in development: More than just noise. *Development (Cambridge, England)*, *136*(23), 3853–3862. https://doi.org/10.1242/dev.035139

Huang, Y., Gao, B.-Q., Meng, Q., Yang, L.-Z., Ma, X.-K., Wu, H., Pan, Y.-H., Yang, L., Li, D., & Chen, L.-L. (2023). CRISPR-dCas13-tracing reveals transcriptional memory and limited mRNA export in developing zebrafish embryos. *Genome Biology*, *24*(1), 15. https://doi.org/10.1186/s13059-023-02848-6

Hwang, B., Lee, J. H., & Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*, *50*(8), Article 8. https://doi.org/10.1038/s12276-018-0071-8

Iyer-Biswas, S., Hayot, F., & Jayaprakash, C. (2009). Stochasticity of gene products from transcriptional pulsing. *Physical Review E*, *79*, 031911.

Johnson, A., Lewis, J., & ALBERTS, B. (2002). *Molecular biology of the cell*.

Jones, D. L., Brewster, R. C., & Phillips, R. (2014). Promoter architecture dictates cell-to-cell variability in gene expression. *Science*, *346*(6216), 1533–1536. https://doi.org/10.1126/science.1255301

Junkin, M., Kaestli, A. J., Cheng, Z., Jordi, C., Albayrak, C., Hoffmann, A., & Tay, S. (2016). High-Content Quantification of Single-Cell Immune Dynamics. *Cell Reports*, *15*(2), 411–422. https://doi.org/10.1016/j.celrep.2016.03.033

Kampen, N. G. (2007). Chapter V - THE MASTER EQUATION. In *Stochastic Processes in Physics and Chemistry* (Third Edition), N.G. Van Kampen, pp. 96–133). Elsevier.

Kempe, H., Schwabe, A., Crémazy, F., Verschure, P. J., & Bruggeman, F. J. (2015). The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Molecular Biology of the Cell*, *26*(4), 797–804. https://doi.org/10.1091/mbc.E14-08-1296

Kierzek, A. M., Zaim, J., & Zielenkiewicz, P. (2001). The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *The Journal of Biological Chemistry*, *276*(11), 8165–8172. https://doi.org/10.1074/jbc.M006264200

Kim, J., & Iyer, V. R. (2004). Global Role of TATA Box-Binding Protein Recruitment to Promoters in Mediating Gene Expression Profiles. *Molecular and Cellular Biology*, *24*(18), 8104–8112. https://doi.org/10.1128/MCB.24.18.8104-8112.2004

Kim, J., & Marioni, J. C. (2013). Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biology*, *14*(1), R7. https://doi.org/10.1186/gb-2013-14-1-r7

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, *220*, 671–680.

Ko, M. S. H. (1991). A stochastic model for gene induction. *Journal of Theoretical Biology*, *153*(2), 181–194. https://doi.org/10.1016/S0022-5193(05)80421-7

Ko, M. S., Nakauchi, H., & Takahashi, N. (1990). The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *The EMBO Journal*, *9*, 2835–2842.

Kramer, Ij. M. (2016). Chapter 13—Activation of the Innate Immune System: The Toll-Like Receptor-4 and Signaling through Ubiquitinylation. In Ij. M. Kramer (Ed.), *Signal Transduction (Third Edition)* (Third Edition, pp. 741–775). Academic Press. https://doi.org/10.1016/B978-0-12-394803-8.00013-9

Kussell, E., & Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, *309*, 2075–2078.

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., Segerstolpe, Å., Rivera, C. M., Ren, B., & Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. *Nature*, *565*(7738), 251–254. https://doi.org/10.1038/s41586-018-0836-1

Lehner, B. (2008). Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Molecular Systems Biology*, *4*, 170. https://doi.org/10.1038/msb.2008.11

Lillacci, G., & Khammash, M. (2010). Parameter Estimation and Model Selection in Computational Biology. *PLOS Computational Biology*, *6*(3), e1000696. https://doi.org/10.1371/journal.pcbi.1000696

Lin, Y. T., & Buchler, N. E. (2019). Exact and efficient hybrid Monte Carlo algorithm for accelerated Bayesian inference of gene expression models from snapshots of single-cell transcripts. *The Journal of Chemical Physics*, *151*(2), 024106. https://doi.org/10.1063/1.5110503

Lionnet, T., & Singer, R. H. (2012). Transcription goes digital. *EMBO reports*, *13*, 313–321.

Little, S. C., Tikhonov, M., & Gregor, T. (2013). Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity. *Cell*, *154*, 789–800.

Liu, T., Zhang, L., Joo, D., & Sun, S.-C. (2017). NF-κB signaling in inflammation. *Signal Transduction and Targeted Therapy*, *2*(1), Article 1. https://doi.org/10.1038/sigtrans.2017.23

Loewer, A., & Lahav, G. (2011). We are all individuals: Causes and consequences of non-genetic heterogeneity in mammalian cells. *Current Opinion in Genetics & Development*, *21*(6), 753–758. https://doi.org/10.1016/j.gde.2011.09.010

Lok, L., & Brent, R. (2005). Automatic generation of cellular reaction networks with Moleculizer 1.0. *Nature Biotechnology*, *23*, 131–136.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, Y., Xue, Q., Eisele, M. R., Sulistijo, E. S., Brower, K., Han, L., Amir, E. D., Pe'er, D., Miller-Jensen, K., & Fan, R. (2015). Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *P Natl Acad Sci USA*, *112*, 607–615.

Lugowski, A., Nicholson, B., & Rissland, O. S. (2018). Determining mRNA half-lives on a transcriptome-wide scale. *Methods*, *137*, 90–98. https://doi.org/10.1016/j.ymeth.2017.12.006

Magklara, A., & Lomvardas, S. (2013). Stochastic gene expression in mammals: Lessons from olfaction. *Trends in Cell Biology*, *23*(9), 449–456. https://doi.org/10.1016/j.tcb.2013.04.005

Maisonneuve, E., & Gerdes, K. (2014). Molecular mechanisms underlying bacterial persisters. *Cell*, *157*, 539–548.

Marshall, J. S., Warrington, R., Watson, W., & Kim, H. L. (2018). An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology*, *14*(2), 49. https://doi.org/10.1186/s13223-018-0278-1

McAdams, H. H., & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(3), 814–819.

McCollum, J. M., Peterson, G. D., Cox, C. D., Simpson, M. L., & Samatova, N. F. (2006). The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior. *Computational Biology and Chemistry*, *30*, 39–49.

McQuarrie, D. A. (1967). Stochastic approach to chemical kinetics. *Journal of Applied Probability*, *4*, 413–478.

Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. *Nature*, *449*(7164), Article 7164. https://doi.org/10.1038/nature06246

Meissner, F., Scheltema, R. A., Mollenkopf, H. J., & Mann, M. (2013). Direct Proteomic Quantification of the Secretome of Activated Immune Cells. *Science*, *340*, 475–478.

Mendes, P., & Kell, D. (1998). Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics*, *14*, 869–883.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, *21*, 1087–1092.

Morgan, M. D., & Marioni, J. C. (2018). CpG island composition differences are a source of gene expression noise indicative of promoter responsiveness. *Genome Biology*, *19*(1), 81. https://doi.org/10.1186/s13059-018-1461-x

Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., & Zimmer, C. (2013). FISH-quant: Automatic counting of transcripts in 3D FISH images. *Nature Methods*, *10*, 277–278.

Muller, K. E. (2001). Computing the confluent hypergeometric function, M(a,b,x). *Numerische Mathematik*, *90*(1), 179–196. https://doi.org/10.1007/s002110100285

Muramoto, T., Cannon, D., Gierliński, M., Corrigan, A., Barton, G. J., & Chubb, J. R. (2012). Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. *Proceedings of the National Academy of Sciences*, *109*, 7350–7355.

Murphy, K. F., Balazsi, G., & Collins, J. J. (2007). Combinatorial promoter design for engineering noisy gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 12726–12731.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100. https://doi.org/10.1016/S0022-2496(02)00028-7

Neuert, G., Munsky, B., Tan, R. Z., Teytelman, L., Khammash, M., & Oudenaarden, A. (2013). Systematic Identification of Signal-Activated Stochastic Gene Regulation. *Science*, *339*, 584–587.

Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., & Weissman, J. S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, *441*, 840–846.

Nicolas, D., Phillips, N., & Naef, F. (2017). *What shapes eukaryotic transcriptional bursting? Mol BioSyst 13*.

Nicolas, D., Zoller, B., Suter, D., & Naef, F. (2018). Modulation of transcriptional burst frequency by histone acetylation. *Proceedings of the National Academy of Sciences*, *115*, 201722330.

Novick, A., & Weiner, M. (1957). ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. *Proceedings of the National Academy of Sciences of the United States of America*, *43*, 553–566.

Oda, K., & Kitano, H. (2006a). A comprehensive map of the toll-like receptor signaling network. *Mol Syst Biol*, *2*, 2006 0015.

Oda, K., & Kitano, H. (2006b). A comprehensive map of the toll-like receptor signaling network. *Molecular Systems Biology*, *2*, 2006.0015. https://doi.org/10.1038/msb4100057

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nature Genetics*, *31*, 69–73.

Padovan-Merhar, O., Nair, G. P., Biaesch, A. G., Mayer, A., Scarfone, S., Foley, S. W., Wu, A. R., Churchman, L. S., Singh, A., & Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global

transcriptional mechanisms. *Molecular Cell*, *58*(2), 339–352. https://doi.org/10.1016/j.molcel.2015.03.005

Park, S. H., Kang, K., Giannopoulou, E., Qiao, Y., Kang, K., Kim, G., Park-Min, K.-H., & Ivashkiv, L. B. (2017). Type I interferons and the cytokine TNF cooperatively reprogram the macrophage epigenome to promote inflammatory activation. *Nature Immunology*, *18*(10), 1104–1116. https://doi.org/10.1038/ni.3818

Pastinen, T., Sladek, R., Gurd, S., Sammak, A., Ge, B., Lepage, P., Lavergne, K., Villeneuve, A., Gaudin, T., Brändström, H., Beck, A., Verner, A., Kingsley, J., Harmsen, E., Labuda, D., Morgan, K., Vohl, M.-C., Naumova, A. K., Sinnett, D., & Hudson, T. J. (2004). A survey of genetic and epigenetic variation affecting human gene expression. *Physiological Genomics*, *16*(2), 184–193. https://doi.org/10.1152/physiolgenomics.00163.2003

Paszek, P. (2007). Modeling Stochasticity in Gene Regulation: Characterization in the Terms of the Underlying Distribution Function. *Bulletin of Mathematical Biology*, *69*, 1567–1601.

Paszek, P., Ryan, S., Ashall, L., Sillitoe, K., Harper, C. V., Spiller, D. G., Rand, D. A., & White, M. R. H. (2010). Population robustness arising from cellular heterogeneity. *Proceedings of the National Academy of Sciences*, *107*(25), 11644–11649. https://doi.org/10.1073/pnas.0913798107

Pawitan, Y. (n.d.). *In All Likelihood*.

Peccoud, J., & Ycart, B. (1995). Markovian Modeling of Gene-Product Synthesis. *Theoretical Population Biology*, *48*(2), 222–234. https://doi.org/10.1006/tpbi.1995.1027

Phillips, N. E., Mandic, A., Omidi, S., Naef, F., & Suter, D. M. (2019). Memory and relatedness of transcriptional activity in mammalian cell lineages. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-09189-8

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., & Beutler, B. (1998). Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: Mutations in Tlr4 gene. *Science (New York, N.Y.)*, *282*(5396), 2085–2088. https://doi.org/10.1126/science.282.5396.2085

Raj, A., Bogaard, P., Rifkin, S. A., Oudenaarden, A., & Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, *5*, 877–879.

Raj, A., & Oudenaarden, A. (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, *135*, 216–226.

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*, *4*(10), e309. https://doi.org/10.1371/journal.pbio.0040309

Raj, A., & van Oudenaarden, A. (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, *135*(2), 216–226. https://doi.org/10.1016/j.cell.2008.09.050

Ramirez-Carrozzi, V. R., Braas, D., Bhatt, D. M., Cheng, C. S., Hong, C., Doty, K. R., Black, J. C., Hoffmann, A., Carey, M., & Smale, S. T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell*, *138*, 114–128.

Ran, D., Zhang, S., Lytal, N., & An, L. (2020). scDoc: Correcting drop-out events in single-cell RNA-seq data. *Bioinformatics*, *36*(15), 4233–4239. https://doi.org/10.1093/bioinformatics/btaa283

Rand, U., Rinas, M., Schwerk, J., Nöhren, G., Linnes, M., Kröger, A., Flossdorf, M., Kály-Kullai, K., Hauser, H., & Höfer, T. (2012). Multi-layered stochasticity and paracrine signal

propagation shape the type-I interferon response. *Molecular Systems Biology*, *8*, 584.

Raser, J. M., & O'Shea, E. K. (2004). Control of Stochasticity in Eukaryotic Gene Expression. *Science*, *304*, 1811–1814.

Raser, J. M., & O'Shea, E. K. (2005). Noise in Gene Expression: Origins, Consequences, and Control. *Science (New York, N.Y.)*, *309*(5743), 2010–2013. https://doi.org/10.1126/science.1105891

Ravarani, C. N. J., Chalancon, G., Breker, M., de Groot, N. S., & Babu, M. M. (2016). Affinity and competition for TBP are molecular determinants of gene expression noise. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms10417

Rhee, H. S., & Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, *483*(7389), 295–301. https://doi.org/10.1038/nature10799

Rivière, I., Sunshine, M. J., & Littman, D. R. (1998). Regulation of IL-4 Expression by Activation of Individual Alleles. *Immunity*, *9*(2), 217–228. https://doi.org/10.1016/S1074-7613(00)80604-4

Robles-Rebollo, I., Cuartero, S., Canellas-Socias, A., Wells, S., Karimi, M. M., Mereu, E., Chivu, A. G., Heyn, H., Whilding, C., Dormann, D., Marguerat, S., Rioja, I., Prinjha, R. K., Stumpf, M. P. H., Fisher, A. G., & Merkenschlager, M. (2022). Cohesin couples transcriptional bursting probabilities of inducible enhancers and promoters. *Nature Communications*, *13*(1), 4342. https://doi.org/10.1038/s41467-022-31192-9

Rodgers, J. R., Johnson, M. L., & Rosen, J. M. (1985). [46] Measurement of mRNA concentration and mRNA half-life as a function of hormonal treatment. In *Methods*

in Enzymology (Vol. 109, pp. 572–592). Elsevier. https://doi.org/10.1016/0076-6879(85)09116-9

Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., & Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science*, *307*, 1962–1965.

Rossi, R. J. (2018). *Mathematical Statistics: An Introduction to Likelihood Based Inference*. Wiley. https://books.google.co.uk/books?id=ehpfDwAAQBAJ

Saint-Antoine, M., Grima, R., & Singh, A. (2022). *A fluctuation-based approach to infer kinetics and topology of cell-state switching* (p. 2022.03.30.486492). bioRxiv. https://doi.org/10.1101/2022.03.30.486492

Sanchez, A., Garcia, H. G., Jones, D., Phillips, R., & Kondev, J. (2011). Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression. *PLoS Computational Biology*, *7*(3), e1001100. https://doi.org/10.1371/journal.pcbi.1001100

Sanchez, A., & Golding, I. (2013). Genetic Determinants and Cellular Constraints in Noisy Gene Expression. *Science*, *342*, 1188–1193.

Sarkar, A., & Stephens, M. (2021). Separating measurement and expression models clarifies confusion in single-cell RNA sequencing analysis. *Nature Genetics*, *53*(6), 770–777. https://doi.org/10.1038/s41588-021-00873-4

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K. D., Imai, T., & Ueda, H. R. (2013). Quartz-Seq: A highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biology*, *14*, 31–31.

Scapigliati, G., Buonocore, F., & Mazzini, M. (2006). Biological Activity of Cytokines: An Evolutionary Perspective. *Current Pharmaceutical Design*, *12*(24), 3071–3081. https://doi.org/10.2174/138161206777947489

Schulte, M. B., & Andino, R. (2014). Single-cell analysis uncovers extensive biological noise in poliovirus replication. *Journal of Virology*, *88*(11), 6205–6212. https://doi.org/10.1128/JVI.03539-13

Schuster, P., & Érdi, J. T. (1989). P. *Berichte der Bunsengesellschaft für physikalische Chemie*, *93*, 1501–1502.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., & Wollman, R. (2014). Accurate information transmission through dynamic biochemical signaling networks. *Science*, *346*, 1370–1373.

Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F. E., Zimmer, C., Mueller, F., & Darzacq, X. (2014). Transcription Factors Modulate c-Fos Transcriptional Bursts. *Cell Reports*, *8*(1), 75–83. https://doi.org/10.1016/j.celrep.2014.05.053

Sepúlveda, L. A., Xu, H., Zhang, J., Wang, M., & Golding, I. (2016). Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science*, *351*, 1218–1222.

Shaffer, S. M., Emert, B. L., Reyes Hueros, R. A., Cote, C., Harmange, G., Schaff, D. L., Sizemore, A. E., Gupte, R., Torre, E., Singh, A., Bassett, D. S., & Raj, A. (2020). Memory Sequencing Reveals Heritable Single-Cell Gene Expression Programs Associated with Distinct Cellular Behaviors. *Cell*, *182*(4), 947-959.e17. https://doi.org/10.1016/j.cell.2020.07.003

Shahrezaei, V., & Swain, P. S. (2008a). Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, *105*, 17256–17261.

Shahrezaei, V., & Swain, P. S. (2008b). Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, *105*(45), 17256–17261. https://doi.org/10.1073/pnas.0803850105

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., Trombetta, J. J., Gennert, D., Gnirke, A., Goren, A., Hacohen, N., Levin, J. Z., Park, H., & Regev, A. (2013a). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, *498*(7453), Article 7453. https://doi.org/10.1038/nature12172

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., Trombetta, J. J., Gennert, D., Gnirke, A., Goren, A., Hacohen, N., Levin, J. Z., Park, H., & Regev, A. (2013b). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, *498*(7453), 236–240. https://doi.org/10.1038/nature12172

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., Chen, P., Gertner, R. S., Gaublomme, J. T., Yosef, N., Schwartz, S., Fowler, B., Weaver, S., Wang, J., Wang, X., Ding, R., Raychowdhury, R., Friedman, N., Hacohen, N., … Regev, A. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, *510*(7505), Article 7505. https://doi.org/10.1038/nature13437

Sharon, E., van Dijk, D., Kalma, Y., Keren, L., Manor, O., Yakhini, Z., & Segal, E. (2014). Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Research*, *24*(10), 1698–1706. https://doi.org/10.1101/gr.168773.113

Silander, O. K., Nikolic, N., Zaslaver, A., Bren, A., Kikoin, I., Alon, U., & Ackermann, M. (2012). A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli. *PLOS Genetics*, *8*, 1002443.

Singer, Z. S., Yong, J., Tischler, J., Hackett, J. A., Altinok, A., Surani, M. A., Cai, L., & Elowitz, M. B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Molecular Cell*, *55*(2), 319–331. https://doi.org/10.1016/j.molcel.2014.06.029

Singh, A., Razooky, B., Cox, C. D., Simpson, M. L., & Weinberger, L. S. (2010). Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophysical Journal*, *98*, 32–34.

Skinner, S. O., Xu, H., Nagarkar-Jaiswal, S., Freire, P. R., Zwaka, T. P., & Golding, I. (2016). Single-cell analysis of transcription kinetics across the cell cycle. *ELife*, *5*, e12175. https://doi.org/10.7554/eLife.12175

Skupsky, R., Burnett, J. C., Foley, J. E., Schaffer, D. V., & Arkin, A. P. (2010). HIV Promoter Integration Site Primarily Modulates Transcriptional Burst Size Rather Than Frequency. *PLOS Computational Biology*, *6*, 1000952.

Smiley, M. W., & Proulx, S. R. (2010). Gene expression dynamics in randomly varying environments. *Journal of Mathematical Biology*, *61*(2), 231–251. https://doi.org/10.1007/s00285-009-0298-z

So, L., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R., & Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. *Nature Genetics*, *43*(6), 554–560. https://doi.org/10.1038/ng.821

Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M., & Sorger, P. K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, *459*(7245), 428–432. https://doi.org/10.1038/nature08012

Spudich, J. L., & Koshland, D. E. (1976). Non-genetic individuality: Chance in the single cell. *Nature*, *262*(5568), Article 5568. https://doi.org/10.1038/262467a0

Srinivas, M., & Patnaik, L. M. (1994). Genetic algorithms: A survey. *Computer*, *27*, 17–26.

Stern, S., Dror, T., Stolovicki, E., Brenner, N., & Braun, E. (2007). Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Molecular Systems Biology*, *3*, 106.

Sturm, G., List, M., & Zhang, J. D. (2021). Tissue heterogeneity is prevalent in gene expression studies. *NAR Genomics and Bioinformatics*, *3*(3), lqab077. https://doi.org/10.1093/nargab/lqab077

Sung, M.-H., Li, N., Lao, Q., Gottschalk, R. A., Hager, G. L., & Fraser, I. D. C. (2014). Switching of the Relative Dominance Between Feedback Mechanisms in Lipopolysaccharide-Induced NF-κB. *Signaling*, *7*, 6–6.

Sung, M.-H., & McNally, J. G. (2011). Live cell imaging and systems biology. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, *3*(2), 167–182. https://doi.org/10.1002/wsbm.108

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., & Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science (New York, N.Y.)*, *332*(6028), 472–474. https://doi.org/10.1126/science.1198817

Svensson, V. (2020). Droplet scRNA-seq is not zero-inflated. *Nature Biotechnology*, *38*(2), 147–150. https://doi.org/10.1038/s41587-019-0379-5

Swain, P. S., Elowitz, M. B., & Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, *99*(20), 12795–12800. https://doi.org/10.1073/pnas.162041399

Symmons, O., & Raj, A. (2016). What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Molecular Cell*, *62*(5), 788–802. https://doi.org/10.1016/j.molcel.2016.05.023

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., & Siddiqui, A. (2009). MRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*, 377–382.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., & Xie, X. S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science (New York, N.Y.)*, *329*(5991), 533–538. https://doi.org/10.1126/science.1188308

Tay, S., Hughey, J. J., Lee, T. K., Lipniacki, T., Quake, S. R., & Covert, M. W. (2010). Single-cell NF-κB dynamics reveal digital activation and analogue information processing. *Nature*, *466*(7303), Article 7303. https://doi.org/10.1038/nature09145

Tirosh, I., & Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. *Genome Research*, *18*, 1084–1091.

Tong, A. J., Liu, X., Thomas, B. J., Lissner, M. M., Baker, M. R., Senagolage, M. D., Allred, A. L., Barish, G. D., & Smale, S. T. (2016). A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. *Cell*, *165*, 165–179.

Tunnacliffe, E., & Chubb, J. R. (2020). What Is a Transcriptional Burst? *Trends in Genetics*, *36*(4), 288–297. https://doi.org/10.1016/j.tig.2020.01.003

Turvey, S. E., & Broide, D. H. (2010). Innate immunity. *The Journal of Allergy and Clinical Immunology*, *125*(2 Suppl 2), S24-32. https://doi.org/10.1016/j.jaci.2009.07.016

Van Eyndhoven, L. C., Verberne, V. P. G., Bouten, C. V. C., Singh, A., & Tel, J. (2023). Transiently heritable fates and quorum sensing drive early IFN-I response dynamics. *ELife*, *12*, e83055. https://doi.org/10.7554/eLife.83055

Verstraeten, N., Knapen, W. J., Kint, C. I., Liebens, V., Bergh, B., Dewachter, L., Michiels, J. E., Fu, Q., David, C. C., & Fierro, A. C. (2015). Obg and Membrane Depolarization Are

Part of a Microbial Bet-Hedging Strategy that Leads to Antibiotic Tolerance. *Molecular Cell*, *59*, 9–21.

Viñuelas, J., Kaneko, G., Coulon, A., Vallin, E., Morin, V., Mejia-Pous, C., Kupiec, J.-J., Beslon, G., & Gandrillon, O. (2013). Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biology*, *11*, 15.

Vu, T. N., Wills, Q. F., Kalari, K. R., Niu, N., Wang, L., Rantalainen, M., & Pawitan, Y. (2016). Beta-Poisson model for single-cell RNA-seq data analyses. *Bioinformatics (Oxford, England)*, *32*(14), 2128–2135. https://doi.org/10.1093/bioinformatics/btw202

Wagner, J. R., Ge, B., Pokholok, D., Gunderson, K. L., Pastinen, T., & Blanchette, M. (2010). Computational analysis of whole-genome differential allelic expression data in human. *PLoS Computational Biology*, *6*(7), e1000849. https://doi.org/10.1371/journal.pcbi.1000849

Waterland, R. A. (2006). Epigenetic mechanisms and gastrointestinal development. *The Journal of Pediatrics*, *149*(5, Supplement), S137–S142. https://doi.org/10.1016/j.jpeds.2006.06.064

Wernet, M. F., Mazzoni, E. O., Çelik, A., Duncan, D. M., Duncan, I., & Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, *440*, 174–180.

Wills, Q. F., Livak, K. J., Tipping, A. J., Enver, T., Goldson, A. J., Sexton, D. W., & Holmes, C. (2013). Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nature Biotechnology*, *31*(8), 748–752. https://doi.org/10.1038/nbt.2642

Wimmers, F., Subedi, N., van Buuringen, N., Heister, D., Vivié, J., Beeren-Reinieren, I., Woestenenk, R., Dolstra, H., Piruska, A., Jacobs, J. F. M., van Oudenaarden, A., Figdor, C. G., Huck, W. T. S., de Vries, I. J. M., & Tel, J. (2018). Single-cell analysis reveals that stochasticity and paracrine signaling control interferon-alpha production by plasmacytoid dendritic cells. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-05784-3

Wu, B., Chao, J. A., & Singer, R. H. (2012). Fluorescence Fluctuation Spectroscopy Enables Quantitative Imaging of Single mRNAs in Living Cells. *Biophysical Journal*, *102*, 2936–2944.

Xue, Q., Lu, Y., Eisele, M. R., Sulistijo, E. S., Khan, N., Fan, R., & Miller-Jensen, K. (2015). *Analysis of single-cell cytokine secretion reveals a role for paracrine signaling in coordinating macrophage responses to TLR4 stimulation*.

Yang, L.-Z., Gao, B.-Q., Huang, Y., Wang, Y., Yang, L., & Chen, L.-L. (2022). Multi-color RNA imaging with CRISPR-Cas13b systems in living cells. *Cell Insight*, *1*(4), 100044. https://doi.org/10.1016/j.cellin.2022.100044

Yang, L.-Z., Wang, Y., Li, S.-Q., Yao, R.-W., Luan, P.-F., Wu, H., Carmichael, G. G., & Chen, L.-L. (2019). Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Molecular Cell*, *76*, 981-997 987.

Yuan, L., Chan, G. C., Beeler, D., Janes, L., Spokes, K. C., Dharaneeswaran, H., Mojiri, A., Adams, W. J., Sciuto, T., Garcia-Cardeña, G., Molema, G., Kang, P. M., Jahroudi, N., Marsden, P. A., Dvorak, A., Regan, E. R., & Aird, W. C. (2016). A role of stochastic phenotype switching in generating mosaic endothelial cell heterogeneity. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms10160

Yunger, S., Rosenfeld, L., Garini, Y., & Shav-Tal, Y. (2010). Single-allele analysis of transcription kinetics in living mammalian cells. *Nature Methods*, *7*(8), 631–633. https://doi.org/10.1038/nmeth.1482

Zenklusen, D., Larson, D. R., & Singer, R. H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural & Molecular Biology*, *15*(12), Article 12. https://doi.org/10.1038/nsmb.1514

Zhang, M. J., Ntranos, V., & Tse, D. (2020). Determining sequencing depth in a single-cell RNA-seq experiment. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-14482-y

Zhao, M., Zhang, J., Phatnani, H., Scheu, S., & Maniatis, T. (2012). Stochastic Expression of the Interferon-β Gene. *PLOS Biology*, *10*, 1001249.

Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A. C., Ghane, T., Di Felice, R., & Rohs, R. (2013). DNAshape: A method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Research*, *41*(Web Server issue), W56-62. https://doi.org/10.1093/nar/gkt437

Zoller, B., Nicolas, D., Molina, N., & Naef, F. (2015). Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology*, *11*, 823.